



```

QY 254 CCTATCCAGTATGAGGGGTGATGATGACATTGAAAGTTTTCGATTACAAAGCCCAAGTA 313
D 259 CCTATCCAGTATGAGGGGTGATGATGACATTGAAAGTTTTCGATTACAAAGCCCAAGTA 318
QY 314 TGGGAATTTAAATGAGNACCTTCCTGGGACATTTGAAAATTTATGATGATGCGCATTTGA 373
D 319 TGGGAATTTAAATGAGNACCTTCCTGGGACATTTGAAAATTTATGATGATGCGCATTTGA 378
QY 374 TTCGAAAGTACATTAACTGGCAAGAACTACACATGGAATGATGAACTTTTCCAACTT 433
D 379 TTCGAAAGTACATTAACTGGCAAGAACTACACATGGAATGATGAACTTTTCCAACTT 438
QY 434 GGGCACTGTAACATTTCCCATCTCCGACCTGAAATGAGATGCCCTTTCTGCTGTAATCAA 493
D 439 GGGCACTGTAACATTTCCCATCTCCGACCTGAAATGAGATGCCCTTTCTGCTGTAATCAA 498
QY 494 GGGCGTCCCTGCTTTTTCGAGGAAATGATGATGTTCACTGGAAGAAAATGGGACATT- 552
D 499 GGGCGTCCCTGCTTTTTCGAGGAAATGATGATGTTCACTGGAAGAAAATGGGACATT- 558
QY 553 AGTTCAAGTAGCAACTATATC-AGGAAACATGTTCAACCAATGAGC-AAAGTGGGTGAAA 610
D 559 AGTTCAAGTAGCAACTATATCAGGAAACATGTTCAACCAATGAGGAAAATGGGTGAAA 618
QY 611 CAGG--ACATGAAACAGGAATTTATTTATGAG--CATGGAATGTAAAAGCCAGCCCA 665
D 619 CAGGCAATGAAACAGGAATTTATTTATGAGAAACATGGAATTTAAAAGCCCAACCAA 678
QY 666 AAAAGGGG 674
D 679 NAAAAANGG 687

```

# RESULT 6 AAK92158

ID AAK92158 standard; cDNA; 697 BP.

AC AAK92158;

DT 06-NOV-2001 (first entry)

DE Human cDNA 5'-end sequence, SEQ ID NO: 618.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

PN EP130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-00114089.

PR 08-JUL-1999; 99JP-00194486.

PR 11-JAN-2000; 2000JP-00118774.

PR 02-MAY-2000; 2000JP-00183765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI; 2001-524255/58.

XX 830 Primers useful for synthesizing full length cDNA clones and their use

PT in genetic manipulation.

XX Claim 2; SEQ ID NO 618; 1380bp + Sequence Listing; English.

XX The invention relates to primers for synthesizing full length cDNA

CC clones. 830 cDNA molecules encoding a human protein have been isolated

CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have

CC been determined. Primers for synthesizing the full length cDNA are useful

for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesized by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is the nucleotide sequence of the 5'-end of a cDNA provided in the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;

Query Match 52.3%; Score 567; DB 4; Length 697;

Best Local Similarity 95.2%; Pred. No. 4.9e-137;

Matches 637; Conservative 0; Mismatches 24; Indels 8; Gaps 5;

```

QY 14 GGGCAGAGATGAGACACGCGCACAGGGCGCGGAGATGCGCGGGGCGCGGCGCTCGG 73
D 19 GGGCAGAGATGAGACACGCGCACAGGGCGCGGAGATGCGCGGGGCGCGGCGCTCGG 78
QY 74 GGAACGCGCTTCTGCTGCTGAGCGCTGCGCTGCTTGGCTGCGCGGTTCGCGGCTGG 133
D 79 GGAACGCGCTTCTGCTGCTGAGCGCTGCGCTGCTTGGCTGCGCGGTTCGCGGCTGG 138
QY 134 TCCCGGCTTCCGGGCACTCCCTTCCGGCGCCCACTGCGCGGTCCTTCAAGGCTTTGAC 193
D 139 TCCCGGCTTCCGGGCACTCCCTTCCGGCGCCCACTGCGCGGTCCTTCAAGGCTTTGAC 198
QY 194 TCCCGTCCAAAACCTGATCCTTATTTGCAAGGTAAGTAACTTTCGTCCAATGCGCTCA 253
D 199 TCCCGTCCAAAACCTGATCCTTATTTGCAAGGTAAGTAACTTTCGTCCAATGCGCTCA 258
QY 254 CCTATCCAGTATGAGGGGTGATGATGACATTGAAAGTTTTCGATTACAAAGCCCAAGTA 313
D 259 CCTATCCAGTATGAGGGGTGATGATGACATTGAAAGTTTTCGATTACAAAGCCCAAGTA 318
QY 314 TGGGAATTTAAATGAGNACCTTCCTGGGACATTTGAAAATTTATGATGATGCGCATTTGA 373
D 319 TGGGAATTTAAATGAGNACCTTCCTGGGACATTTGAAAATTTATGATGATGCGCATTTGA 378
QY 374 TTCGAAAGTACATTAACTGGCAAGAACTACACATGGAATGATGAACTTTTCCAACTT 433
D 379 TTCGAAAGTACATTAACTGGCAAGAACTACACATGGAATGATGAACTTTTCCAACTT 438
QY 434 GGCACACTGTAACATTTCCCATCTCCGACCTGAAATGAGATGCCCTTTCTGCTGTAATCAA 493
D 439 GGCACACTGTAACATTTCCCATCTCCGACCTGAAATGAGATGCCCTTTCTGCTGTAATCAA 498
QY 494 GGGCGTCCCTGCTTTTTCGAGGAAATGATGATGTTCACTGGAAGAAAATGGGACATT- 552
D 499 GGGCGTCCCTGCTTTTTCGAGGAAATGATGATGTTCACTGGAAGAAAATGGGACATT- 558
QY 553 AGTTCAAGTAGCAACTATATC-AGGAAACATGTTCAACCAATGAGC-AAAGTGGGTGAAA 610
D 559 AGTTCAAGTAGCAACTATATCAGGAAACATGTTCAACCAATGAGGAAAATGGGTGAAA 618
QY 611 CAGG--ACATGAAACAGGAATTTATTTATGAG--CATGGAATGTAAAAGCCAGCCCA 665
D 619 CAGGCAATGAAACAGGAATTTATTTATGAGAAACATGGAATTTAAAAGCCCAACCAA 678
QY 666 AAAAGGGG 674
D 679 NAAAAANGG 687

```

## RESULT 7 AAI25096

ID AAI25096 standard; DNA; 494 BP.

AC AAI25096;

DT 12-OCT-2001 (first entry)

XX Probe #15029 for gene expression analysis in human cervical cell sample.



CM Probe: human; microarray; gene expression; cervical epithelial cell;  
CM cervical cancer; ss.  
CX Homo sapiens.  
CX WO200157278-A2.  
CN 09-AUG-2001.  
XD 30-JAN-2001; 2001WO-US000670.  
XP 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
KX (MOLE-) MOLECULAR DYNAMICS INC.  
KX Penn SG, Hanzel DK, Chen W, Rank DR;  
KX WPI; 2001-488901/53.  
KX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human cervical epithelial cells.  
KX Claim 25; SEQ ID NO 15029; 487bp; English.  
PS The present invention relates to human single exon nucleic acid probes  
CC (SENPs). The present sequence is one such probe. The SENPs are derived  
CC from human HeLa cells. The SENPs can be used to produce a single exon  
CC microarray, which can be used for measuring human gene expression in a  
CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging of  
CC diseases of the cervix, notably cervical cancer. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
KX  
KX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
SQ  
Query Match 45.3%; Score 491; DB 4; Length 494;  
Best Local Similarity 100.0%; Pred. No. 2,4e-117;  
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 594 TGGCAAGTGGTGAACAGAGCAATGAAACAGAAATTTATGAGACATGGAATGTA 653  
DB 1 TGGCAAGTGGTGAACAGAGCAATGAAACAGAAATTTATGAGACATGGAATGTA 60  
QY 654 AAGCCAGCCAGAAAAGGGGGCAGAGCATGTTGATTCCTAGACATGTTCCAAATTTG 713  
DB 61 AAGCCAGCCAGAAAAGGGGGCAGAGCATGTTGATTCCTAGACATGTTCCAAATTTG 120  
QY 714 TGTTAGACCTTTAACAAGTTGGCTGAATTTGAGCAGAGTTCAAGAACTTAAGAACCA 773  
DB 121 TGTTAGACCTTTAACAAGTTGGCTGAATTTGAGCAGAGTTCAAGAACTTAAGAACCA 180  
QY 774 ACTATACAGAAATATTTCTTTACAGTGGAGAACCACTTATCTGGGAATGAAACATCTG 833  
DB 181 ACTATACAGAAATATTTCTTTACAGTGGAGAACCACTTATCTGGGAATGAAACATCTG 240  
QY 834 TTTTGGGCCAACAGAAACAGACTCTTGGTTTAGCCATAAAAAGATTTTATACCCCT 893  
DB 241 TTTTGGGCCAACAGAAACAGACTCTTGGTTTAGCCATAAAAAGATTTTATACCCCT 300  
QY 894 TCAACCCATTTGGCACTAAAGATTTCTTGAAGTCTCTTGCAATTTTGTATGAG 953  
DB 301 TCAACCCATTTGGCACTAAAGATTTCTTGAAGTCTCTTGCAATTTTGTATGAG 360  
QY 954 TGATTGTGCAAAACAGATTCTATTGTTTAAATTTTGAAATTTGGTTTACCTATGA 1013

DB 361 TGATTGTGCAAAACAGCTTATTTGTTTATTAATTTGTAATTTGGTTTTCACCTATGA 420  
QY 1014 AATTCCTTTTATTAATAATACATATGAGAAATCCCTTACTATCAGAAACAAACAC 1073  
DB 421 AATTCCTTTTATTAATAATACATATGAGAAATCCCTTACTATCAGAAACAAACAC 480  
QY 1074 TCTCTGTTTA 1084  
DB 481 TCTCTGTTTA 491  
RESULT 8  
ABR70786  
ID ABR70786 standard; DNA; 494 BP.  
XX  
XX ABR70786;  
AC  
XX 01-FEB-2002 (first entry)  
DT  
XX Human foetal liver single exon nucleic acid probe #19091.  
DE  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
KX  
KX Homo sapiens.  
OS  
XX WO200157277-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US000669.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
KX (MOLE-) MOLECULAR DYNAMICS INC.  
KX Penn SG, Hanzel DK, Chen W, Rank DR;  
KX WPI; 2001-483447/52.  
KX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human fetal liver.  
KX Claim 4; SEQ ID NO 19091; 639bp + Sequence Listing; English.  
PS The invention relates to a single exon nucleic acid probe for measuring  
CC human gene expression in a sample derived from human foetal liver. The  
CC single exon nucleic acid probes may be used for predicting, measuring and  
CC displaying gene expression in samples derived from human fetal liver. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
KX  
KX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
SQ  
Query Match 45.3%; Score 491; DB 4; Length 494;  
Best Local Similarity 100.0%; Pred. No. 2,4e-117;  
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 594 TGGCAAGTGGTGAACAGAGCAATGAAACAGAAATTTATGAGACATGGAATGTA 653  
DB 1 TGGCAAGTGGTGAACAGAGCAATGAAACAGAAATTTATGAGACATGGAATGTA 60  
QY 654 AAGCCAGCCAGAAAAGGGGGCAGAGCATGTTGATTCCTAGACATGTTCCAAATTTG 713  
DB 61 AAGCCAGCCAGAAAAGGGGGCAGAGCATGTTGATTCCTAGACATGTTCCAAATTTG 120

```

QY 714 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAATAGAAACCA 773
DB 121 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAATAGAAACCA 180
QY 774 ACTATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 833
DB 181 ACTATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 240
QY 834 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 893
DB 241 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 300
QY 894 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGTCTCTGCAATTTTGAATGACG 953
DB 301 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGTCTCTGCAATTTTGAATGACG 360
QY 954 TGAATGTGCAACAACAGTCTATTTGTTTATATTTTGAATTTGTTTACCTATGA 1013
DB 361 TGAATGTGCAACAACAGTCTATTTGTTTATATTTTGAATTTGTTTACCTATGA 420
QY 1014 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 1073
DB 421 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 480
QY 1074 TCTCTGGTTTA 1084
DB 481 TCTCTGGTTTA 491

RESULT 9
AA150961
ID AA150961 standard; DNA; 494 BP.
AC AA150961;
XX
DT 17-OCT-2001 (first entry)
DE Probe #19647 used to measure gene expression in human placenta sample.
KM Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR,
XX
PN WPI; 2001-488897/53.
XX
DR
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
PS Claim 25; SEQ ID NO 19647; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene

```

```

CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders
XX
SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Query Match 45.3%; Score 491; DB 4; Length 494;
Beet Local Similarity 100.0%; Pred. No. 2,4e-117;
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 524 TGGCAAAATGGGTTGAACAAGACAATGAACAAGAAATTAATATGACATGAAATGTA 653
DB 1 TGGCAAAATGGGTTGAACAAGACAATGAACAAGAAATTAATATGACATGAAATGTA 60
QY 654 AAGCCAGCCAGAAAAGGGGCGACAGACATGTTGATTCCTACGACTGTTCCAAATTTG 713
DB 61 AAGCCAGCCAGAAAAGGGGCGACAGACATGTTGATTCCTACGACTGTTCCAAATTTG 120
QY 714 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAATAGAAACCA 773
DB 121 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGGACAGAGTTCAGAAATAGAAACCA 180
QY 774 ACTATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 833
DB 181 ACTATCAAGAAATATTTCTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTG 240
QY 834 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 893
DB 241 TTTTGGGCGCAACAGAAACAAGACTCTGGTTAGCCATAAAAGATTTTACCCCT 300
QY 894 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGTCTCTGCAATTTTGAATGACG 953
DB 301 TCAACCAATTTGGCACTAAAGAAATTTCTGTAGTCTCTGCAATTTTGAATGACG 360
QY 954 TGAATGTGCAACAACAGTCTATTTGTTTATATTTTGAATTTGTTTACCTATGA 1013
DB 361 TGAATGTGCAACAACAGTCTATTTGTTTATATTTTGAATTTGTTTACCTATGA 420
QY 1014 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 1073
DB 421 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTATCAGAAACAAACAC 480
QY 1074 TCTCTGGTTTA 1084
DB 481 TCTCTGGTTTA 491

RESULT 10
ABA37272
ID ABA37272 standard; DNA; 494 BP.
AC ABA37272;
XX
DT 23-JAN-2002 (first entry)
DE Probe #15738 for gene expression analysis in human heart cell sample.
XX
KM Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KM congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.

```

27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488899/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts.  
XX  
PS Claim 4; SEQ ID NO 15738; 530bp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart. The  
XX present sequence is one such probe. The probes may be used for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from the human heart via microarrays. By measuring gene expression, the  
XX probes are useful for predicting, diagnosing, grading, staging, and  
XX monitoring and prognosing diseases of the human heart and vascular system  
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
XX congenital heart disease. Note: The sequence data for this patent did not  
XX form part of the printed specification, but was obtained in electronic  
XX format directly from WPI at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
XX  
Query Match 45.3%; Score 491; DB 4; Length 494;  
Best Local Similarity 100.0%; Pred. No. 2,4e-117;  
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 594 TGGCAAAAGTGGGTGAACAGACCAATGAAACAGAAATTATTATGACATGATGTAA 653  
Db 1 TGGCAAAAGTGGGTGAACAGACCAATGAAACAGAAATTATTATGACATGATGTAA 60  
XX  
QY 654 AAGCCAGCCCAAGAAAGGGGCGAGAGCATGTTGATTCCTAGACGTTCCAAATTG 713  
Db 61 AAGCCAGCCCAAGAAAGGGGCGAGAGCATGTTGATTCCTAGACGTTCCAAATTG 120  
XX  
QY 714 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 773  
Db 121 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 180  
XX  
QY 774 ACTATACAGAATATTTCTTACAGTGGAGAACCTTATCTGAGAAATGAAACATCTG 813  
Db 181 ACTATACAGAATATTTCTTACAGTGGAGAACCTTATCTGAGAAATGAAACATCTG 240  
XX  
QY 813 TTTTGGGCGCAACAGAAACAAGACTCTTGGTTAGCCATATAAAGATTTTATTAACCCCT 893  
Db 241 TTTTGGGCGCAACAGAAACAAGACTCTTGGTTAGCCATATAAAGATTTTATTAACCCCT 300  
XX  
QY 894 TCAAACCAATTTGGCACTAAGAAATTCCTGTAAGTCTCTGCAATTTTGTATGAG 953  
Db 301 TCAAACCAATTTGGCACTAAGAAATTCCTGTAAGTCTCTGCAATTTTGTATGAG 360  
XX  
QY 954 TGAATGTCACAACAAGTTCTAATTTGTTTAAATTTGAAATTTGTTTACCTATGA 1013  
Db 361 TGAATGTCACAACAAGTTCTAATTTGTTTAAATTTGAAATTTGTTTACCTATGA 420  
XX  
QY 1014 AATTCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1073  
Db 421 AATTCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480  
XX  
QY 1074 TCTCTGTTTA 1084  
Db 481 TCTCTGTTTA 491  
XX  
RESULT 11  
AAK44992  
ID AAK44992 standard; DNA; 494 BP.  
XX

AAK44992;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human bone marrow expressed single exon probe SEQ ID NO: 19549.  
XX  
XX Human bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukemia; lymphoma; myeloma; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157276-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000668.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human bone marrow.  
XX  
XX Example 4; SEQ ID NO 19549; 658bp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukemia and myeloma. The present sequence is one of  
XX the probes of the invention  
XX  
XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
XX  
Query Match 45.3%; Score 491; DB 4; Length 494;  
Best Local Similarity 100.0%; Pred. No. 2,4e-117;  
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 594 TGGCAAAAGTGGGTGAACAGACCAATGAAACAGAAATTATTATGACATGATGTAA 653  
Db 1 TGGCAAAAGTGGGTGAACAGACCAATGAAACAGAAATTATTATGACATGATGTAA 60  
XX  
QY 654 AAGCCAGCCCAAGAAAGGGGCGAGAGCATGTTGATTCCTAGACGTTCCAAATTG 713  
Db 61 AAGCCAGCCCAAGAAAGGGGCGAGAGCATGTTGATTCCTAGACGTTCCAAATTG 120  
XX  
QY 714 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 773  
Db 121 TGTTAAGACCTTTAAACAAGTTGGCTGAATTGGAGCAGAGTTCAAGAACATAGAACCA 180  
XX  
QY 774 ACTATACAGAATATTTCTTACAGTGGAGAACCTTATCTGAGAAATGAAACATCTG 813  
Db 181 ACTATACAGAATATTTCTTACAGTGGAGAACCTTATCTGAGAAATGAAACATCTG 240  
XX  
QY 813 TTTTGGGCGCAACAGAAACAAGACTCTTGGTTAGCCATATAAAGATTTTATTAACCCCT 893  
Db 241 TTTTGGGCGCAACAGAAACAAGACTCTTGGTTAGCCATATAAAGATTTTATTAACCCCT 300  
XX  
QY 894 TCAAACCAATTTGGCACTAAGAAATTCCTGTAAGTCTCTGCAATTTTGTATGAG 953  
Db 301 TCAAACCAATTTGGCACTAAGAAATTCCTGTAAGTCTCTGCAATTTTGTATGAG 360  
XX

```
QY 954 TGATTGTGCAGAAACAGTCTATTGTTTATTAATTTGAATNGGTTTAACTATGA 1013
DB 361 TGAATTGTGCAGAAACAGTCTATTGTTTATTAATTTGAATNGGTTTAACTATGA 420
QY 1014 AATTCCCTTTTATTAATAATACATATGAGAAATCCCTTTACTATCAGAAACAAACAC 1073
DB 421 AATTCCCTTTTATTAATAATACATATGAGAAATCCCTTTACTATCAGAAACAAACAC 480
QY 1074 TCTCTGGTTTA 1084
DB 481 TCTCTGGTTTA 491

RESULT 12
AAK19040
ID AAK19040 standard; DNA; 494 BP.
XX
XX AAK19040;
AC
XX
XX 05-NOV-2001 (first entry)
DE Human brain expressed single exon probe SEQ ID NO: 19031.
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
XX ss.
XX Homo sapiens.
XX
XX MO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US000667.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO 19031; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is one of the probes of the
XX invention
XX
XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
XX
XX Query Match 45.3%; Score 491; DB 4; Length 494;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-117;
XX Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 594 TGGCAAAAGTGGTGAACACAGACATGAACAGAAATTTATTTGAGACATGGAATGTA 653
DB 1 TGGCAAAAGTGGTGAACACAGACATGAACAGAAATTTATTTGAGACATGGAATGTA 60
QY 654 AAGCCAGCCCAAGAAAAGGGGCGAGACATGTTGATTCTTACGACTGTTCAAATTGG 713
```

```
DB 61 AAGCCAGCCCAAGAAAAGGGGCGAGACATGTTGATTCTTACGACTGTTCAAATTGG 120
QY 714 TGTAAAGACCTTTAAACAAGTTGCTGAATTTGAGCAAGTTCAAGAACATAGAACCA 773
DB 121 TGTAAAGACCTTTAAACAAGTTGCTGAATTTGAGCAAGTTCAAGAACATAGAACCA 180
QY 774 ACTATACAGAAATATTTCTTTACAGTGAAGAACTTACTTATCTGGAAATGAAACATCTG 833
DB 181 ACTATACAGAAATATTTCTTTACAGTGAAGAACTTACTTATCTGGAAATGAAACATCTG 240
QY 834 TTTTGGGCGCAACAGAAACAAAGACTCTGTGTTAGCCATAAAAGATTTTATACCCT 893
DB 241 TTTTGGGCGCAACAGAAACAAAGACTCTGTGTTAGCCATAAAAGATTTTATACCCT 300
QY 894 TCAACCAATTTGCACTAAGAAATTTCTGTAGTCTCTGCAAAATTTTGAATGACAG 953
DB 301 TCAACCAATTTGCACTAAGAAATTTCTGTAGTCTCTGCAAAATTTTGAATGACAG 360
QY 954 TGATTGTGCAGAAACAGTCTATTGTTTATTAATTTGAATNGGTTTAACTATGA 1013
DB 361 TGAATTGTGCAGAAACAGTCTATTGTTTATTAATTTGAATNGGTTTAACTATGA 420
QY 1014 AATTCCCTTTTATTAATAATACATATGAGAAATCCCTTTACTATCAGAAACAAACAC 1073
DB 421 AATTCCCTTTTATTAATAATACATATGAGAAATCCCTTTACTATCAGAAACAAACAC 480
QY 1074 TCTCTGGTTTA 1084
DB 481 TCTCTGGTTTA 491

RESULT 13
ABS44658
ID ABS44658 standard; DNA; 494 BP.
XX
XX ABS44658;
AC
XX
XX 25-FEB-2003 (first entry)
DE Human liver single exon probe, SEQ ID NO 19648.
XX
XX Human; single exon nucleic acid probe; liver; cirrhosis;
XX hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
XX coronary heart disease; ss.
XX
XX Homo sapiens.
XX
XX MO200157273-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488898/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX
XX Claim 4; SEQ ID NO 19648; 658bp; English.
XX
```

CC The invention relates to a single exon nucleic acid probe (SENP) (1) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (1) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABS25011-ABS51005 represent human  
CC liver single exon nucleic acid probes of the invention. Note: The  
CC sequence information for this patent does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX  
XX Sequence 494 BP, 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
SQ

Query Match 45.3%; Score 491; DB 4; Length 494;  
Best Local Similarity 100.0%; Pred. No. 2,4e-117;  
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DY 594 TGGCAAGTGGGTGAAACAGACCAATGAAACAGAAATTATTATGACATGAAATGTA 653  
DB 1 TGGCAAGTGGGTGAAACAGACCAATGAAACAGAAATTATTATGACATGAAATGTA 60  
DY 654 AAGCAGCCGAGAAAGGGGGGAGAGACATGTTGATCTTACGACTGTTCCAAATTTG 7.3  
DB 61 AAGCAGCCGAGAAAGGGGGGAGAGACATGTTGATCTTACGACTGTTCCAAATTTG 120  
DY 714 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGAGCAGAGTTGACAGAAACATAGAAACA 773  
DB 121 TGTAAAGACCTTTAAACAAGTTGGCTGAATTTGAGCAGAGTTGACAGAAACATAGAAACA 180  
DY 774 ACTATACAGAAATATTTCTTACAGTGAGAACCTACTTATCTGGGAAATGAAACATCTG 833  
DB 181 ACTATACAGAAATATTTCTTACAGTGAGAACCTACTTATCTGGGAAATGAAACATCTG 240  
DY 834 TTTTGGGCGCAACGAAACAGAAAGCTGTTGTTAGCCATAAAGATTTATTAACCCCT 893  
DB 241 TTTTGGGCGCAACGAAACAGAAAGCTGTTGTTAGCCATAAAGATTTATTAACCCCT 300  
DY 894 TCAACACCATTTGCCAATGAAAGATTTCTGTTAGCTCTTCCAAATTTTGTATGAG 953  
DB 301 TCAACACCATTTGCCAATGAAAGATTTCTGTTAGCTCTTCCAAATTTTGTATGAG 360  
DY 954 TGATTGTGCAAAACAGTTCTATTGTTTAAATTTGAAATATTTGTTTACCTATGA 1013  
DB 361 TGATTGTGCAAAACAGTTCTATTGTTTAAATTTGAAATATTTGTTTACCTATGA 420  
DY 1014 AATTCCCTTTATTAATAATACATATGAAAGATCCCTTACCTATGCAAGAAACAC 1073  
DB 421 AATTCCCTTTATTAATAATACATATGAAAGATCCCTTACCTATGCAAGAAACAC 480  
DY 1074 TCTCTGCTTTA 1084  
DB 481 TCTCTGCTTTA 491

RESULT 14  
ABS19237  
ID ABS19237 standard; DNA; 494 BP.  
XX  
XX ABS19237;  
AC  
AC  
DT 19-AUG-2002 (first entry)  
XX  
XX Human genome-derived single exon probe ORF from lung SEQ ID No 19228.  
DE  
XX  
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease; open reading frame; ORF.  
OS  
OS Homo sapiens.  
XX  
XX WO200186003-A2.  
PD  
PD 15-NOV-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000665.  
PF  
PF 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI  
PI WPI; 2002-114183/15.  
XX  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.  
XX  
XX Claim 4; SEQ ID NO 19228; 634bp; English.  
XX  
XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung; comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a single exon probe open reading frame of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 494 BP, 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
SQ

Query Match 45.3%; Score 491; DB 6; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 2,4e-117;  
 Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 594 TGGCAAGGGGTGAACAAGCAATGAACAAGCAATTTATATAGACATGAGATGTA 653
DB 1 TGGCAAGGGGTGAACAAGCAATGAACAAGCAATTTATATAGACATGAGATGTA 60
QY 654 AAGCCAGCCCAAGAAAAGGGGCGAGAGACATGGTTGATTCCTACACAGTTCCCAATTTG 713
DB 61 AAGCCAGCCCAAGAAAAGGGGCGAGAGACATGGTTGATTCCTACACAGTTCCCAATTTG 120
QY 714 TGTAAAGACCTTTAAACAAGTTGCTGAATTTGAGCAGAGTTCAAGAACATAGAACCA 773
DB 121 TGTAAAGACCTTTAAACAAGTTGCTGAATTTGAGCAGAGTTCAAGAACATAGAACCA 190
QY 774 ACTATACAAAGATATTTCTTTACATGAGAACCTTATCTGGAATGGAATGAAACATG 833
DB 181 ACTATACAAAGATATTTCTTTACATGAGAACCTTATCTGGAATGGAATGAAACATG 240
QY 834 TTTTGGGCGAACAAGAAACAAGACTCTGTTGTTAGCATTAAGATTTTATACCCCT 893
DB 241 TTTTGGGCGAACAAGAAACAAGACTCTGTTGTTAGCATTAAGATTTTATACCCCT 300
QY 894 TCAAAACACATTTGCCAATTAAGAAATTTCTGTAGCTCTTGCATAATTTTGATGAG 953
DB 301 TCAAAACACATTTGCCAATTAAGAAATTTCTGTAGCTCTTGCATAATTTTGATGAG 360
QY 954 TCAATGTGCAACAAGCTCTATTTGTTTATATTTGAATTTGATTTTACCTATGA 1013
DB 361 TCAATGTGCAACAAGCTCTATTTGTTTATATTTGAATTTGATTTTACCTATGA 420
QY 1014 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTACAGAAACAAACAC 1073
DB 421 AATTCCTTTTATTAATAATACATATGAAGAAATCCCTTACCTACAGAAACAAACAC 480
QY 1074 TCTCTGCTTTA 1084
DB 481 TCTCTGCTTTA 491

RESULT 15
AAV88347
ID AAV88347 standard; cDNA; 506 BP.
XX
AC AAV88347;
XX
DT 12-FEB-1999 (first entry)
XX
DE EST clone GB814.
XX
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemokines; chemokines; haemostasis; gene therapy; chromobolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
XX Homo sapiens.
XX
XX OS
XX PN MO9845437-A2.
XX
XX PD 15-OCT-1998.
XX
XX PF 10-APR-1998; 98WO-US006956.
XX
XX PR 10-APR-1997; 97US-00837312.
XX
XX PA (GENE) GENETICS INST INC.
XX
XX PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Werberg D, Treacy M;
XX PI Spaulding V, Agostino MJ;
XX DR WPI; 1999-070078/06.

```

FT New polynucleotides encoding human secreted proteins - derived from e.g.  
 PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,  
 PI pituitary, retina and colon cDNA libraries.

XX Claim 1; Page 366; 641pp; English.

CC The present sequence represents an expressed sequence tag (EST), and is a  
 CC polynucleotide of the invention. The polynucleotides of the invention are  
 CC all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene therapy

XX SQ Sequence 506 BP; 171 A; 86 C; 119 G; 130 T; 0 U; 0 Other;

Query Match 37.0%; Score 401.4; DB 2; Length 506;  
 Best Local Similarity 99.8%; Pred. No. 4.7e-94;  
 Matches 402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 497 GCTGCTGCTTTTGAAGGAATGATGATGTCATGGAAGAAATGAGACATTAGTT 556
DB 99 GCTGCTGCTTTTGAAGGAATGATGATGTCATGGAAGAAATGAGACATTAGTT 158
QY 557 CAAGTAGCACTATATCAGAAACATGTTCAACCAATGCGAAAGTGGGTGAACAGAC 616
DB 159 CAAGTAGCACTATATCAGAAACATGTTCAACCAATGCGAAAGTGGGTGAACAGAC 218
QY 617 AATGAACAGAAATTTATATGAGACATGAGATGTAAGCCAGCCGAAAGGGGGCA 676
DB 219 AATGAACAGAAATTTATATGAGACATGAGATGTAAGCCAGCCGAAAGGGGGCA 278
QY 677 GAGACATGTTGATTCCTACAGCTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 736
DB 279 GAGACATGTTGATTCCTACAGCTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 338
QY 737 GCTGAATTTGAGACAGAGTTTAAGACATAGAAACCAACTATACAAAGATTTCTTAC 796
DB 339 GCTGAATTTGAGACAGAGTTTAAGACATAGAAACCAACTATACAAAGATTTCTTAC 398
QY 797 AGTGAGAACCTACTATCTGGAAGAAATGAAACATCTGTTTGGGCGAAGAGAAACAG 856
DB 399 AGTGAGAACCTACTATCTGGAAGAAATGAAACATCTGTTTGGGCGAAGAGAAACAG 458
QY 857 ACTCTGTTTGAAGCAATTAAGATTTTATTAACCCCTTCAAC 899
DB 459 ACTCTGTTTGAAGCAATTAAGATTTTATTAACCCCTTCAAC 501

```

Search completed: April 22, 2004, 00:39:08  
 Job time : 487.252 secs





COMMENT

ZYMOGENETICS INC

OS Homo sapiens (human)

PN JP 2001511345-A/1

PD 14-AUG-2001

PF 24-JUL-1998 JP 2000504249

PR 24-JUL-1997 US 60/053613

PI PAUL O SHEPPARD, DIERA G GILBERTON

PC C12N15/09, A61K38/00, A61K48/00, C07K14/47, C07K16/18, C12N1/15, PC C12N1/19, PC

C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12N15/00, A61K37/02, C12N5/PC

CC Secreted protein which human chromosome 13 encodes FH Key

FEATURES

FT CDS Location/Qualifiers

source 1..1486

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1084; DB 6; Length 1486;

Best Local Similarity 100.0%; Pred. No. 1,1e-238;

Matches 1084; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GAATTCGGCAGCGAGGCGAGAGGTAGACACGCGCAGCGCGCCGAGATCGCGGCGGC 60

1 GAATTCGGCAGCGAGGCGAGAGGTAGACACGCGCAGCGCGCCGAGATCGCGGCGGC 60

61 GGGCGCGGCTCGGGAGCGCGCTTCTGCTGCTGGCGCTGCGCGCTTTGGCTCGCGGT 120

61 GGGCGCGGCTCGGGAGCGCGCTTCTGCTGCTGGCGCTGCGCGCTTTGGCTCGCGGT 120

121 GGTTCGGGCTGGTCCCGGCTCTCGGGCACTCCCTCCCGGCGCCACTGGCGGCTCCCTA 180

121 GGTTCGGGCTGGTCCCGGCTCTCGGGCACTCCCTCCCGGCGCCACTGGCGGCTCCCTA 180

181 CAAAGGCTTTGACTCCGTCGCAAAACCTGATCCTATTGTCAAGTAGTAATCTTTCTG 240

181 CAAAGGCTTTGACTCCGTCGCAAAACCTGATCCTATTGTCAAGTAGTAATCTTTCTG 240

241 TCCAACTGGCTCACTATCCCACTTATGAGAGGCTGATGACATTTGAAGTTTTCGATT 300

241 TCCAACTGGCTCACTATCCCACTTATGAGAGGCTGATGACATTTGAAGTTTTCGATT 300

301 ACAAGCCAGTAGTGGAACTTAATATGAGAGACCTCTGGGACACTTGAATTTATGCA 360

301 ACAAGCCAGTAGTGGAACTTAATATGAGAGACCTCTGGGACACTTGAATTTATGCA 360

361 TGATGCCATTGGATTCAAGAGTACATTAACTGGCAAGAACTACACAAATGATGTA 420

361 TGATGCCATTGGATTCAAGAGTACATTAACTGGCAAGAACTACACAAATGATGTA 420

421 ACTTTTCCAACTGGCAACTGTACATTTCCCATCTCGACCTGAAATGATGCCCTTT 480

421 ACTTTTCCAACTGGCAACTGTACATTTCCCATCTCGACCTGAAATGATGCCCTTT 480

481 CTGGTATATCAAGGCGCTGCTCTTTTGAAGGAAATGATGATTCATCTGGAAGA 540

481 CTGGTATATCAAGGCGCTGCTCTTTTGAAGGAAATGATGATTCATCTGGAAGA 540

541 AAATGGGACATTAGTTCAAGTAGCAACTATATCAGAGAAACATGTTCAACCAATGGCAA 600

541 AAATGGGACATTAGTTCAAGTAGCAACTATATCAGAGAAACATGTTCAACCAATGGCAA 600

601 GTGGGTGAACAGGCAATGAAACGAAATTTATATGAGACATGAAATGTAAGCCAG 660

601 GTGGGTGAACAGGCAATGAAACGAAATTTATATGAGACATGAAATGTAAGCCAG 660

661 CCCAGAAAAGGGGCGAGAGCATGGTTGATTCCTACGACCTGTTCCAAATTTGTGTTAG 720

661 CCCAGAAAAGGGGCGAGAGCATGGTTGATTCCTACGACCTGTTCCAAATTTGTGTTAG 720

QY 721 GACCTTTAACAAGTTGCTGATTTGGAGCAGAGTTCAAGACATAGAAAACAATATAC 780

DB 721 GACCTTTAACAAGTTGCTGATTTGGAGCAGAGTTCAAGACATAGAAAACAATATAC 780

QY 781 AAGAAATTTCTTTTACGTGAGAGAACTTATCTTATCTGGGAAATGAAACATCTGTTTTGG 840

DB 781 AAGAAATTTCTTTTACGTGAGAGAACTTATCTTATCTGGGAAATGAAACATCTGTTTTGG 840

QY 841 GCCAAGCAGAAAGAGCTGTTTACGATTAAGATTAAGATTTATACCTCCCTCAAAAC 900

DB 841 GCCAAGCAGAAAGAGCTGTTTACGATTAAGATTAAGATTTATACCTCCCTCAAAAC 900

QY 901 ACATTTGCCAATTAAGAAATTTCTGTTAGAGTCTCTTCAAAATTTTATGAGAGATTTGT 960

DB 901 ACATTTGCCAATTAAGAAATTTCTGTTAGAGTCTCTTCAAAATTTTATGAGAGATTTGT 960

QY 961 GCAAAACAGTTCTATTGTTTATTAATTTGAATTTGTTTACCTATGAATTTCCC 1020

DB 961 GCAAAACAGTTCTATTGTTTATTAATTTGAATTTGTTTACCTATGAATTTCCC 1020

QY 1021 TTTTATTAATAATACATATGAGAAATCCCTTACCTATCAGAAACAAACACCTCTCG 1080

DB 1021 TTTTATTAATAATACATATGAGAAATCCCTTACCTATCAGAAACAAACACCTCTCG 1080

QY 1081 TTTTA 1084

DB 1081 TTTTA 1084

RESULT 2

AF068227

LOCUS 4080 bp mRNA linear PRI 28-JUL-1998

DEFINITION Homo sapiens putative transmembrane protein (CLNS) mRNA, complete cds.

ACCESSION AF068227

VERSION AF068227.1 GI:3342385

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 4080)

Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.

CLNS, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis

Nat. Genet. 19 (3), 286-288 (1998)

JOURNAL

MEDLINE

PUBMED

9662406

2 (bases 1 to 4080)

Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.

Direct Submission

Submitted (26-MAY-1998) Department of Human Molecular Genetics, National Public Health Institute, Mannerheimintie 166, Helsinki 00300, Finland

FEATURES

source

1..4080

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/chromosome="13"

/map="13q22"

1..4080

/gene="CLNS"

1293..2516

/gene="CLNS"

/codon\_start=1

/product="putative transmembrane protein"

/protein\_id="AAC27614.1"

/db\_xref="GI:3342386"



Matches 1067; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

QY 14 GGGCAGAGGTAGACACGCGCACAGGCGCGCGAGATCGCGCGCGCGCGCGCGCGCG 73
   |
Db 16 GGGCAGAGGTAGACACGCGCACAGGCGCGCGAGATCGCGCGCGCGCGCGCGCGCG 75
   |
QY 74 GGAACGCGCTTCTCGTGTGCTGGGCGCTGCGCGCTGCTTGGCTCGCGGTGTTCCGGCTGG 133
   |
Db 76 GGAACGCGCTTCTCGTGTGCTGGGCGCTGCGCGCTGCTTGGCTCGCGGTGTTCCGGCTGG 135
   |
QY 134 TCCCGGGTCTCGGGGCTTCCCGTCCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGC 193
   |
Db 136 TCCCGGGTCTCGGGGCTTCCCGTCCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGC 195
   |
QY 194 TTCCCGGCAAAAAGCTGATCTTATTTGCTCAAGCTAAGTACTTCTGCTCAACTGGCTCA 253
   |
Db 196 TTCCCGGCAAAAAGCTGATCTTATTTGCTCAAGCTAAGTACTTCTGCTCAACTGGCTCA 255
   |
QY 254 CCTATCCCAAGTTATGAGAGGCTGATGATGACATTTGATTTTGGATTACAGCCCGAGTA 313
   |
Db 256 CCTATCCCAAGTTATGAGAGGCTGATGATGACATTTGATTTTGGATTACAGCCCGAGTA 315
   |
QY 314 TGGGAATTAATTAATGAGAACCTTCCCGGAGCACTTGAATAATTAATGATGATGATGATG 373
   |
Db 316 TGGGAATTAATTAATGAGAACCTTCCCGGAGCACTTGAATAATTAATGATGATGATGATG 375
   |
QY 374 TTGAGAGTACATTAATCTGCAAGAACTACACATGAGATGAGATGAGATGAGATGAGATG 433
   |
Db 376 TTGAGAGTACATTAATCTGCAAGAACTACACATGAGATGAGATGAGATGAGATGAGATG 435
   |
QY 434 GGGCACTGTACATTTCCCACTCCGACCTCGAATGAGATGAGATGAGATGAGATGAGATG 493
   |
Db 436 GGGCACTGTACATTTCCCACTCCGACCTCGAATGAGATGAGATGAGATGAGATGAGATG 495
   |
QY 494 GGGCGCGCTGCTTTTGAAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 553
   |
Db 496 GGGCGCGCTGCTTTTGAAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 555
   |
QY 554 GTTCAAGTACAACTATATCAGGAAACATGTTCAACCAATGCAAAAGTGGTGAACAG 613
   |
Db 556 GTTCAAGTACAACTATATCAGGAAACATGTTCAACCAATGCAAAAGTGGTGAACAG 615
   |
QY 614 GACAAATGAAACAGAAATTTATATGAGACATGATGATGATGATGATGATGATGATGATG 673
   |
Db 616 GACAAATGAAACAGAAATTTATATGAGACATGATGATGATGATGATGATGATGATGATG 675
   |
QY 674 GAGAGACATGATTTGATCTTACGACCTGTTCCAAATTTGTTTAAAGACCTTTTAAACAG 733
   |
Db 676 GAGAGACATGATTTGATCTTACGACCTGTTCCAAATTTGTTTAAAGACCTTTTAAACAG 735
   |
QY 734 TTGGCTGAATTTGAGACAGTTCAAGACATAGAAACCACTATACAGAAATATTTCTT 793
   |
Db 736 TTGGCTGAATTTGAGACAGTTCAAGACATAGAAACCACTATACAGAAATATTTCTT 795
   |
QY 794 TACAGTGAAGAACCTTATCTTGTGGAATGAAACCTTGTGTTTGGGCAACAGAAAC 853
   |
Db 796 TACAGTGAAGAACCTTATCTTGTGGAATGAAACCTTGTGTTTGGGCAACAGAAAC 855
   |
QY 854 AAGACTCTGGTTAGCAATAAAGATTTTATACCCCTCAACCAACATTTGGCCAACT 913
   |
Db 856 AAGACTCTGGTTAGCAATAAAGATTTTATACCCCTCAACCAACATTTGGCCAACT 915
   |
QY 914 AAGAAATTTCTGTAGTCTCTTCAAAATTTTGTAGTGAAGTGTGTCAGAAACAGTTC 973
   |
Db 916 AAGAAATTTCTGTAGTCTCTTCAAAATTTTGTAGTGAAGTGTGTCAGAAACAGTTC 975
   |
QY 974 TATTTGTTTAAATTTGATGATTTGTTTATACCTATGAAATCCCTTTATTAATAATA 1033
   |
Db 976 TATTTGTTTAAATTTGATGATTTGTTTATACCTATGAAATCCCTTTATTAATAATA 1035
   |
QY 1034 ACATATGAAGAAATCCCTTACCTATCAGAAACAAACAACTCTCTGGTTA 1084
   |
Db 1036 ACATATGAAGAAATCCCTTACCTATCAGAAACAAACAACTCTCTGGTTA 1086

```

RESULT 4  
BD127858 2120 bp DNA linear PAT 18-SEP-2002  
LOCUS BD127858  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION BD127858  
VERSION BD127858.1 GI:23222803  
KEYWORDS JP 2002017375-A/3289.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.  
1 (bases 1 to 2120)  
TITLE Primer for synthesizing full-length cDNA and use thereof  
JOURNAL Patent: JP 2002017375-A 3289 22-JAN-2002;  
COMMENT HELIX RESEARCH INSTITUTE  
OS Homo sapiens (human)  
PN JP 2002017375-A/3289  
PD 22-JAN-2002  
PF 07-UTB-2000 JP 2000253172  
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO PI  
PI ISHII,  
PI YURI KAWAI, AI MAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI SHINICHI KOJIMA,  
PI TETSUJI OTSUKI, HISASHI KOGA  
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
FT CDS Location/Qualifiers  
1..2120  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 98.2%; Score 1064.6; DB 6; Length 2120;  
Best Local Similarity 99.6%; Pred. No. 3.5e-234;  
Matches 1067; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
14 GGGCAGAGGTAGACACGCGCACAGGCGCGCGAGATCGCGCGCGCGCGCGCGCGCG 73  
19 GGGCAGAGGTAGACACGCGCACAGGCGCGCGAGATCGCGCGCGCGCGCGCGCGCG 78  
74 GGAACGCGCTTCTCGTGTGCTGGGCGCTGCGCGCTGCTTGGCTCGCGGTGTTCCGGCTGG 133  
79 GGAACGCGCTTCTCGTGTGCTGGGCGCTGCGCGCTGCTTGGCTCGCGGTGTTCCGGCTGG 138  
134 TCCCGGGTCTCGGGGCTTCCCGTCCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGC 193  
139 TCCCGGGTCTCGGGGCTTCCCGTCCCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTGGC 198  
194 TTCCCGGCAAAAAGCTGATCTTATTTGCTCAAGCTAAGTACTTCTGCTCAACTGGCTCA 253  
199 TTCCCGGCAAAAAGCTGATCTTATTTGCTCAAGCTAAGTACTTCTGCTCAACTGGCTCA 258  
254 CCTATCCCAAGTTATGAGAGGCTGATGATGACATTTGATTTTGGATTACAGCCCGAGTA 313  
259 CCTATCCCAAGTTATGAGAGGCTGATGATGACATTTGATTTTGGATTACAGCCCGAGTA 318  
314 TGGGAATTAATTAATGAGAACCTTCCCGGAGCACTTGAATAATTAATGATGATGATGATG 373  
319 TGGGAATTAATTAATGAGAACCTTCCCGGAGCACTTGAATAATTAATGATGATGATGATG 378  
374 TTGAGAGTACATTAATCTGCAAGAACTACACATGAGATGAGATGAGATGAGATGAGATG 433

Db	379	TTGAGAACTACATTAACCTGGCAAGAACTACACAAATGGAATGATATGAACCTTTCCAACTT	438
Qy	434	GGCAACTGTACATTTTCCCATCTCCGACCTGTAATGATGCCCTTTCTGGTGTATCA	493
Db	439	GGCAACTGTACATTTTCCCATCTCCGACCTGTAATGATGCCCTTTCTGGTGTATCA	498
Qy	494	GGGCTGCGCTCTTTTGGAGGAAATGGATATGTCATCTGGAAGGAAATGGACATTA	553
Db	499	GGGCTGCGCTCTTTTGGAGGAAATGGATATGTCATCTGGAAGGAAATGGACATTA	558
Qy	554	GTTCAAGTAGCACTATATCAGAAACATGTTCACCAAAATGGCAAAATGGGTGAAACAG	613
Db	559	GTTCAAGTAGCACTATATCAGAAACATGTTCACCAAAATGGCAAAATGGGTGAAACAG	618
Qy	614	GACAAATGAAACAGAAATTTATTTATGACATGAAATGTAATAAGCCAGCCAGAAAAGGG	673
Db	619	GACAAATGAAACAGAAATTTATTTATGACATGAAATGTAATAAGCCAGCCAGAAAAGGG	678
Qy	674	GCAAGACATGGTTTGAATTCCTACGACCTGTCCAAATTTGGTTAAGAACCTTTAACAG	733
Db	679	GCAAGACATGGTTTGAATTCCTACGACCTGTCCAAATTTGGTTAAGAACCTTTAACAG	738
Qy	734	TTGGCTGAATTTGGAGCAGAGTTGACAGAACTAGAAACCACTATACAGAAATTTTCTT	793
Db	739	TTGGCTGAATTTGGAGCAGAGTTGACAGAACTAGAAACCACTATACAGAAATTTTCTT	798
Qy	794	TACAGTGAGAACCTACTTATCTGGGAATGAAACATCTGTTTTTGGGCCAACAGAAAC	853
Db	799	TACAGTGAGAACCTACTTATCTGGGAATGAAACATCTGTTTTTGGGCCAACAGAAAC	858
Qy	854	AAGACTCTGGTTTGAAGCATTAATAAGATTTTATTAACCCCTTCAAAACCACTTTGGCACT	913
Db	859	AAGACTCTGGTTTGAAGCATTAATAAGATTTTATTAACCCCTTCAAAACCACTTTGGCACT	918
Qy	914	AAAGAAATTTCTGTGAGCTCTTCGCAAAATTTTATGATGAGATTTGGCACAACAGTTC	973
Db	919	AAAGAAATTTCTGTGAGCTCTTCGCAAAATTTTATGATGAGATTTGGCACAACAGTTC	978
Qy	974	TATTTGTTTATATATTTTGAATATTTGGTTTATACCTATGAAATTCCTTTATTTAAATA	1033
Db	979	TATTTGTTTATATATTTTGAATATTTGGTTTATACCTATGAAATTCCTTTATTTAAATA	1038
Qy	1034	ACATATGAGAAATCCCTTTACCTATCAGAAACAAACACCTCTCTGGTTA	1084
Db	1039	ACATATGAGAAATCCCTTTACCTATCAGAAACAAACACCTCTCTGGTTA	1089
RESULT 5			
AK075109			
LOCUS			
DEFINITION	AK075109	2120 bp mRNA linear	FRI 03-SEP-2002
ACCESSION		Homo sapiens cDNA FLJ90628 fis, clone PLACE1003407, highly similar	
VERSION		to Homo sapiens putative transmembrane protein (CIN5) mRNA.	
KEYWORDS		AK075109	
SOURCE		AK075109.1 GI:22760983	
ORGANISM		Oligo capping; Eris (full insert sequence).	
REFERENCE		Homo sapiens	
AUTHORS		Ekharoye; Metzosa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
		Isoagai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T.,	
		Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S.,	
		Kawai-Hito, Y., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y.,	
		Kojima, S., Nagahari, K., Masuko, Y., Ono, T., Okano, K., Yoshikawa, Y.,	
		Acetuka, S., Saeki, N., Hattori, A., Okumura, K., Iwayanagi, T. and	
		Ninomiya, K.	
		NEDO human cDNA sequencing project	
		Unpublished	
		2 (bases 1 to 2120)	
		Isoagai, T. and Otsuki, T.	
		Direct Submission	
JOURNAL		Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,	
REFERENCE		Genomics Laboratory, 1532-3-Tana, Kistatazu, Chiba 292-0812, Japan	
AUTHORS			
TITLE			
JOURNAL			

[illegible]

Db 739 TTGGCTGAATTTGAGCAGAGTTCAAGAACATAGAAACCAACTATACAAATATTTCTT 798  
Qy 794 TACAGTGGAGAACTACTATATCGGGGAAATGAACATCTGTTTGGGGCAACAGGAAC 853  
Db 799 TACAGTGGAGAACTACTATATCGGGGAAATGAACATCTGTTTGGGGCAACAGGAAC 858  
Qy 854 AAGACTCTGTTTGGAGCCATAAAGATTTTATTAACCCCTTCAACCACTTTGGCAACT 913  
Db 859 AAGACTCTGTTTGGAGCCATAAAGATTTTATTAACCCCTTCAACCACTTTGGCAACT 918  
Qy 914 AAGAAATTTCTGTGAGTCTCTTGCAGAAATTTTGTATGATGATGTCAGCAACAGTTC 973  
Db 919 AAGAAATTTCTGTGAGTCTCTTGCAGAAATTTTGTATGATGATGTCAGCAACAGTTC 978  
Qy 974 TATTTGTTTATTAATTTGATATTTGTTTATTAACCAATTTGCTTTTATTAATA 1033  
Db 979 TATTTGTTTATTAATTTGATATTTGTTTATTAACCAATTTGCTTTTATTAATA 1038  
Qy 1034 ACATATGAAGAAATCCCTTTACCTATCAAGAAACAAACACTCTCTGTGTTTA 1084  
Db 1039 ACATATGAAGAAATCCCTTTACCTATCAAGAAACAAACACTCTCTGTGTTTA 1089

## RESULT 6

LOCUS BD073413 1038 bp DNA linear PAT 27-AUG-2002  
DEFINITION Secreted protein which human chromosome 13 encodes.

ACCESSION BD073413  
VERSION BD073413.1 GI:22619016

KEYWORDS JP 200151345-A/12.  
SYNTHETIC CONSTRUCT

SOURCE artificial sequences.  
ORGANISM 1 (bases 1 to 1038)

REFERENCE Shepard, P.O. and Gilberton, D.G.  
TITLE Secreted protein which human chromosome 13 encodes  
JOURNAL Patent: JP 200151345-A 12 14-AUG-2001;  
ZYMOGENETICS INC

COMMENT OS Artificial Sequence  
PN JP 200151345-A/12  
PD 14-AUG-2001  
PE 24-JUL-1998 JP 2000504249  
PR 24-JUL-1997 US 60/053613  
PI PAUL, O SHEPPARD, DIBRA G GILBERTON  
PC C12N15/09, A61K38/00, A61K48/00, C07K14/47, C07K16/18, C12N1/15, PC  
C12N1/19,  
PC C12N1/21, C12N5/10, C12P21/02, C12Q1/68, C12N15/00, A61K37/02, C12N5/ PC  
00  
CC Degenerate nucleotide sequence encoding zsig46 polypeptide of  
CC SEQ ID NO:2  
FH Key location/Qualifiers  
FT source 1..1038  
FT location/Qualifiers  
FT location/Qualifiers

FEATURES  
source 1..1038  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

## ORIGIN

Query Match 71.3%; Score 772.4; DB 6; Length 1038;  
Best Local Similarity 61.7%; Pred. No. 5, 2e-167;  
Matches 640; Conservative 232; Mismatches 165; Indels 0; Gaps 0;  
Qy 47 ATGCGCGGCGCGCGCGCGCTGCGGACGCGCTTCCGTGCTGCGCGCTGCGCTG 106  
Db 1 ATGCGCGGCGCGCGCGCGCGCTGCGGACGCGCTTCCGTGCTGCGCGCTGCGCTG 60  
Qy 107 CTTTGGCTGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 166  
Db 61 YNTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120

Qy 167 TGGCGGTGCGCTCAAGCGCTTGAAGTCCGTCGCAAAAGCTGATCTTATGTCAAGCT 226  
Db 121 TGGCGGTGCGCTCAAGCGCTTGAAGTCCGTCGCAAAAGCTGATCTTATGTCAAGCT 180  
Qy 227 AAGTATCTTCTGTGCAAGTGGCTCACTTATCCAGTATGAGGAGTATGATGACAT 286  
Db 181 AAGTATCTTCTGTGCAAGTGGCTCACTTATCCAGTATGAGGAGTATGATGACAT 240  
Qy 287 GAACTTTTGCATTCAGAGCCCGAGTATGGAATTTAATATGAGAGCTCTGGGACAC 346  
Db 241 GAACTTTTGCATTCAGAGCCCGAGTATGGAATTTAATATGAGAGCTCTGGGACAC 300  
Qy 347 TTGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406  
Db 301 TTGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
Qy 407 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 466  
Db 361 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
Qy 467 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 526  
Db 421 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
Qy 527 GTTCACTGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGAT 586  
Db 481 GTTCACTGGAAGAAATGAGCACTTATGATGATGATGATGATGATGATGATGATGAT 540  
Qy 587 AACCAATGCAAAAGTGGTGAAGAGCAATGAAACAGAAATTTATATGAGCATGG 646  
Db 541 AACCAATGCAAAAGTGGTGAAGAGCAATGAAACAGAAATTTATATGAGCATGG 600  
Qy 647 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 706  
Db 601 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 660  
Qy 707 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 766  
Db 661 AATGTAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 720  
Qy 767 GAAACCACTATACAGATATTTCTTTACAGTGAAGTGAAGTGAAGTGAAGTGAAG 826  
Db 721 GAAACCACTATACAGATATTTCTTTACAGTGAAGTGAAGTGAAGTGAAGTGAAG 780  
Qy 827 ACATCTGTTTGGGCGCAACGAAACAGCTCTGTTTGAAGCAATTAAGATTTAT 886  
Db 781 ACATCTGTTTGGGCGCAACGAAACAGCTCTGTTTGAAGCAATTAAGATTTAT 840  
Qy 887 TACCCCTTCAAAACCAATTTGCGCACTTAAAGATTTGTTGAGTCTCTGCAATTTT 946  
Db 841 TACCCCTTCAAAACCAATTTGCGCACTTAAAGATTTGTTGAGTCTCTGCAATTTT 900  
Qy 947 GATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006  
Db 901 GATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
Qy 1007 CCTTGAATATCCCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1066  
Db 961 CCTTGAATATCCCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020

Qy 1067 AAAACACTCTGCTGTTT 1083  
Db 1021 AAAACACTCTGCTGTTT 1037

RESULT 7  
LOCUS BC025487 2318 bp mRNA linear ROD 16-APR-2003  
DEFINITION Mus musculus ceroid- $\beta$ -lipofuscinosis, neuronal 5, mRNA (cDNA clone  
IMAGE:5251891), partial cds.  
ACCESSION BC025487  
VERSION BC025487.1 GI:19343790  
KEYWORDS





Qy	811	TTATCTGGGAATGAAACATCTGTTTGGGCAACAGAAACAGACTCTGCTTATGC	870
Db	737	TTACTGGGAATGAAACATCTATTTTGGGCCAAAGAAACAGACTCTGCTTGGC	796
Qy	871	CATAAAAAGATTTATTAACCCCTTCAACACATTTGCCAATAAGAAATTTCTGTAG	930
Db	797	CATAAAAAATTTATGCCCCCTTCAGACCGATTTGTCAACCAAGATTTTCTATGAA	856
Qy	931	TCTCTGCAATTTTGTATGACGATTTGTGCAACACAGTCTATTGTTTATATTT	990
Db	857	TTTCTGAAAAATTTTGTATGATGATTAATACACAGACAGTTCTACTGTTTAACTT	916
Qy	991	TGAATATTGTTTTTAACTATGAATTCCTTTTATTAATATACATATGAAGAAATCC	1056
Db	917	TGAGATATGTTTCTACCAATGAACCCCCCTTTCTCAAAATTAACATATGAAGAAACCC	976
Qy	1051	TTTACCTATCGAAACAAACACCTCTCGGTTT	1083
Db	977	GTATCTTACCGACATACACATTTTACGACTT	1009

RESULT	8
AC109554/c	
LOCUS	AC109554
DEFINITION	Rattus norvegicus clone CH230-33ID3, *** SEQUENCING IN PROGRESS
ACCESSION	AC109554
VERSION	AC109554.4
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus

REFERENCES

1 (bases 1 to 198524)

Mundy,D,Merrie, Metzker,M, Lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,H, Alpbrooks,S, Amin,A, Angliano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Bismato,K, Blair,T, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E, Cardenas,V, Carter,C, Cavazos,I, Ceasar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,C, Cleeland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,S, Davy-Carroll,L, De Anna,C, Dedrick,D, Delgado,O, Denison,C, Deramo,C, Ding,Y, Dim,H, Ditya,K, Diaper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duvall,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,W, Foster,P, Fraser,C,M, Gabisi,A, Galla,R, Garcia,A, Garner,T, Garza,M, Geregorgios,E, Geer,K, Giller,R, Grady,M, Guerin,W, Guevara,W, Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K, Harvey,J, Hawlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Huliy,S, Hune,U, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolyet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovarc,J, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,D, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lotenshewa,L, Louissegh,H, Lozado,R,U, Lu,X, Ma,T, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawhinny,S, McLeod,M,P, McNell,T,Z, Meenen,E, Milosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundasa,M, Murphy,M, Nair,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Parks,K, Nwackelemel,O, Okwunonu,G, Olarnunsegun,A, Pal,S, Parks,K, Paeternack,S, Paul,H, Perez,A, Perez,L, Pfankoch,C, Plopper,F, Poindexter,A, Popovic,D, Primus,E, Pu,L,-L, Puzio,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reich,R, Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,U, Sanders,W, Savery,G, Scherer,S, Scott,G, Shattman,S, Shen,H,

REFERENCE	JOURNAL TITLE	AUTHORS
1	Direct Submission	Unpublished
2	(bases 1 to 198524)	Worley, K.C.
3	Direct Submission	Submitted (05-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4	(bases 1 to 198524)	Rat Genome Sequencing Consortium.
5	Direct Submission	Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
6	On Oct 11, 2002 this sequence version replaced gi:21738217, version 1.	

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GOLY
Center clone name: CH230-33203
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 177481 bases at least Q40
Consensus quality: 179640 bases at least Q30
Consensus quality: 181141 bases at least Q20
Estimated insert size: 183954; sum-of-coverage estimation
Quality coverage: 6x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1 198524: contig of 198524 bp in length.
Location/Qualifiers
1. .198524

```

misc\_feature

```
misc_feature
/clone_end=T7" _extension
131405..131443
/note="clone_boundary
clone_end:T7
site:Mbol
end_sequence:RXAPC14TV"
197370..198217
/note="clone_boundary
clone_end:Sp6
site:Mbol
end_sequence:RXAPC14TV"
ORIGIN
```

	Query Match	Best Local	Similarity	71.1%;	Score	662.8;	DB	2;	Length	198524;
	Matches	800;	Conservative	0;	Mismatches	242;	Indels	0;	Gaps	0;
2Y	30	CGGCAACAGGCGCCGAGATGCGGCGGCGCGCGGCGCGGCTCGGGACGCGCTTCCTGCT	85							
Db	38682	CGACGCGCTTGCGCCCGGGAGCCACCGGGACCTTCAGGGAGTCCGCGGGCCAGGTGAGCGCC	38623							
2Y	90	GCAGGCGCGTGGCGCTGCTTTGGCTCGCGGTGTTCGCGGCTGTGTCCCGGGTCTCGGGCA	149							
Db	38622	ACTGGGCGCAGGCGGCTCGCGCTGGCGCTACTGGGCTGTGCGGCGACTTCGTGGCGCGCTCC	38563							
2Y	150	TCCCTCCCGGCGCCACTGCGCGGTGCCCTACAGGCGCTTTGACTTCGCTCCAAAACCTG	209							
Db	38562	CGACGCTGGGCAACCGCTGGCCCGTACCTTACAGAGCGCTTCTTCCTTCGCTCGGAGACAG	38503							
2Y	210	ATCCTTATGTCAAGCTAAGTATACTTTCTGTCCACTGGCTACCTTATCCGATTATGG	269							
Db	38502	ATCCCTTCTGTCAAGCCAGATATACCTTTGTGCTTACTGTCTCACCCATCCCATAGTATTA	38443							
2Y	270	AGGATATGATGACATTTGAAGTTTTCATATTCACAGCCCGAGTATGGGAATTTAAATATG	329							
Db	38442	AGGACATATGACGTCAATCGAAGTTTAAAGCTTACAGCCCTGTGTGGGAATTTAAATATG	38383							
2Y	330	GAGACTCTCGGGACACTTGAATAATATGCAATGATGCACTTGATTCAGAGTACATTAA	389							
Db	38382	GAGACTCTCTGGGACACTTCAAAATCATGACAGTCCATTTGATTCAGAGTACGCTGA	38323							
2Y	390	CTGGCAAGAACTACACCAATGGAATGGTATGAAGTTTTCACACTTGGCACTGTACATTTC	449							
Db	38322	CGGCAAGAACTACCAATTTGAATGGTATGAAGTTTTCAGAGCTCGGCACTGTACATTTC	38263							
2Y	450	CCCATCTCCGACCTGAATGATGCCCCCTTCTGTGTATCAAGGCGCTGCTCTCTT	509							
Db	38262	CCCATCTCGGCGCGAAGTGAAGCGTCCGTTCTGTGTATCAAGGCGGCGAGCTCTCTCT	38203							
2Y	510	TTGAGGGAATTGATGATGTTTCACTGGAAGAAAATGGGCACTTATGTTCAATGACACTA	569							
Db	38202	TTGAAAGAAATTTACGATTAACACTGGAAAGAAAACGGGCGTTGTCTGTGTTGCAACCA	38143							
2Y	570	TATCAGGAACAATGTTCAACCAAAATGGCAATGGGTGAAACAGGCAATGAAACAGGA	629							
Db	38142	TATCCGGAACAACGTTTACAAAGTGGCCGAGTGGGTGAAGCAGACATATGAGACTGGGA	38083							
2Y	630	TTTATTTATGACATGGAATGTATAAGCCAGCCAGAAAAGGGGCGACAGACATGCTTTG	689							
Db	38082	TTTATTTACAGACATGGAACGTCGCGGCGCAGGCCAGAAAAGGGGCGACAGACATGCTTTG	38023							
2Y	690	ATTCTCAAGACGTTTCAAAATTTGTGTTTAAAGACCTTTAAACAAGTTGGCTGAATTTGGAG	749							
Db	38022	AATCTCAACACTGTTTCGAATTTTGTCTTTAAGGACATATGAAATTTGGCTGAATTTGAA	37963							
2Y	750	CAGAGTTCAAGAACATAGAAACCAACTATACAAAGATATTTCTTTACAGTGAAGAACTTA	809							
Db	37962	CAGAAATTCAGGAAGTAAAGAAACAACTATATGAATAATTTCTTTACGCGGAGAAACGA	37903							
2Y	810	CTTATCTGGGAATGAACAATCTGTTTGTGGGCCAAGGAAACAAAGACTTTGGTTTATG	869							
Db	37902	TTTATCTGGGAATGAACAATCTTATTTTGTGGGCCCAAGGAAACAAAGACTTTGGTTTATG	37843							

QY	870	CCATTAAGAAAGATTTTATTTTACCCCTTCAAAACACATTTGGCAACTAAAGATTTCTGTGA	929
Db	37842	CCATTAAGAAAGATTTTACGGCCCCCGTCMAACGGTATTTCTGACCAAGATTTTCTGTGA	37783
QY	930	GTCCTCGCAATTTTGTGATGAGTATTTGTCACAAACAGTTCTATTGTTTAAAT	989
Db	37782	ATTCTCTGAAATTTTGTGACAGAGTATTAACAGAGAGTTTCACTGTTTAACT	37723
QY	990	TTCGAATTTGTTTATTTTACCTATGAAATTCCTTTATTAATTAACATATGAAGAAATCC	1049
Db	37722	TTGAGTATTTGTTTCTTACCTATGAACACCCCTTTGTCAAAATTAACATAGAGGAAACCC	37663
QY	1050	CTTTACTCTATACGAACCAAAACACCTCTCG	1079
Db	37662	CTTTACTCTATACCAACATACGACATTTTACC	37633
RESULT 9			
AC107510			
LOCUS			
DEFINITION	AC107510	227920 bp	DNA linear HTG 13-MAY-2003
			Rattus norvegicus clone CH23-13806, *** SEQUENCING IN PROGRESS
			***, 4 unordered pieces.
ACCESSION	AC107510		
VERSION	AC107510.5	GI:30580095	
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Bukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE			
AUTHORS	1 (bases 1 to 227920)		
	Muzny, D., Matle, M., Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angilano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaru, D., Barber, M., Barnes, M., Barnstead, M., Benahmed, F., Biwaldo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, J., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Guneratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Huliy, S., Hume, J., Ididbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Khan, Z., King, L., Kovac, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, D., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshaw, L., Louissege, H., Lopez, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, K., Martinez, E., Mawhinney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minj, E., Montemagno, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nunyan, N., Norris, S., Nwackeleneb, O., Okunolu, G., Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindeexter, A., Popovic, D., Primm, E., Pu, L., L., Puzos, B., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojars, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Sma, S. D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sos, J., Stemle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,		



Db 38102 ATTCTTGAAATTTTGGACACAGTATGACACAGAGGTTCTACCTGTTTATTAACCT 38161  
QY 990 TTGAATATTTGGTTTTTACCTATGAAATTCCTTTATTAATAAACA 1036  
Db 38162 TTGAGATTGGTTTCTTACCTATGAAACCCCTTTGCTGCTAGGA 38208

RESULT 10  
BD125187 697 bp DNA linear PAT 18-SEP-2002  
LOCUS BD125187  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION BD125187  
VERSION BD125187.1 GI:23220132  
KEYWORDS JP 2002017375-A/618.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 697)  
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,  
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and  
Koga,H.  
Primer for synthesizing full-length cDNA and use thereof  
Patent: JP 2002017375-A 618 22-JAN-2002;  
HELIX RESEARCH INSTITUTE  
OS Homo sapiens (human)  
PN JP 2002017375-A/618  
PD 22-JAN-2002  
PF 07-JUL-2000 JP 2000253172  
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO  
PI ISHII,  
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI  
SHINICHI KOJIMA,  
PI TETSUO OTSUKI, HISASHI KOGA  
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10,  
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
FT source 1..697  
Location/Qualifiers  
/organism="Homo sapiens (human)".  
1..697  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 52.3%; Score 567; DB 6; Length 697;  
Best Local Similarity 95.2%; Pred. No. 8,9e-120;  
Matches 637; Conservative 0; Mismatches 24; Indels 8; Gaps 5;

QY 14 GGGCAGAGGTAGACACGCGACAGCGCGCGAGATGCGCGCGGCGCGCGGCTCGG 73  
Db 19 GCGCAGAGGTAGACACGCGACAGCGCGCGAGATGCGCGCGGCGCGGCTCGG 78  
QY 74 GGAGCGGCTTCTGCTGCTGGCGCTGGCGCTGCTTGGCTCGCGGTGTTCCGGCTGG 133  
Db 79 GGAGCGGCTTCTGCTGCTGGCGCTGGCGCTGCTTGGCTCGCGGTGTTCCGGCTGG 138  
QY 134 TCCGCGGCTGCTGGCGATCCCTCCCGGCGCACTGCGCGGCTGCAAGAGCTTTGAC 193  
Db 139 TCCGCGGCTGCTGGCGATCCCTCCCGGCGCACTGCGCGGCTGCAAGAGCTTTGAC 198  
QY 194 TTCCGCTCAAACTGATCTTATTTGCTAAGCTAAGTACTTTTGTCTCAACTGGCTCA 253  
Db 199 TTCCGCTCAAACTGATCTTATTTGCTAAGCTAAGTACTTTTGTCTCAACTGGCTCA 258  
QY 254 CCTATCCAGTTATGAGGGGTGATGATGACATGTAAGTTTTCATTAACAAGCCCACTA 323  
Db 259 CCTATCCAGTTATGAGGGGTGATGATGACATGTAAGTTTTCATTAACAAGCCCACTA 328

QY 314 TGGGAATTTAAATATGAGACCTCTCGGACACTTGAATAATTATGATGATGCCATTGA 373  
Db 319 TGGGAATTTAAATATGAGACCTCTCTCGGACACTTGAATAATTATGATGATGCCATTGA 378  
QY 374 TTGGAAGTACATTACTGCGCAAGACTACCAATGGAATGATGAACTTTTCAACTT 433  
Db 379 TTGGAAGTACATTACTGCGCAAGACTACCAATGGAATGATGAACTTTTCAACTT 438  
QY 434 GCGAAGTACATTACTTCCCATCTCCGACCTGGAATGGAATGCCCTTTCTGATGATCA 493  
Db 439 GCGAAGTACATTACTTCCCATCTCCGACCTGGAATGGAATGCCCTTTCTGATGATCA 498  
QY 494 GCGCCTCGCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGAT 552  
Db 499 GCGCCTCGCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGAT 558  
QY 553 AGTTCAAGTACCAATATATC-AGGAAACATGTTCAACCAATATGCG-AAAAGTGGTGA 610  
Db 559 AGTTCAAGTACCAATATATCAAGAAACATGTTCAACCAATATGCGCAAAAGTGGTGA 618  
QY 611 CAGG--ACATGAAACAGGAATTTATTTAGAG--CATGAAATGTAAGACCCAGCCAG 665  
Db 619 CAAAGCAATGAAACAGGAATTTATTTAGAGCAATGAAATTTAAAGCCANCCCA 678  
QY 666 AAAAGGGG 674  
Db 679 NAAAAANG 687

RESULT 11  
BD126453 697 bp DNA linear PAT 18-SEP-2002  
LOCUS BD126453  
DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
ACCESSION BD126453  
VERSION BD126453.1 GI:23221398  
KEYWORDS JP 2002017375-A/1884.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 697)  
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,  
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and  
Koga,H.  
Primer for synthesizing full-length cDNA and use thereof  
Patent: JP 2002017375-A 1884 22-JAN-2002;  
HELIX RESEARCH INSTITUTE  
OS Homo sapiens (human)  
PN JP 2002017375-A/1884  
PD 22-JAN-2002  
PF 07-JUL-2000 JP 2000253172  
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO  
PI ISHII,  
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI  
SHINICHI KOJIMA,  
PI TETSUO OTSUKI, HISASHI KOGA  
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC  
10,  
PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
Primer for synthesizing full-length cDNA and use thereof FH Key  
FT source 1..697  
Location/Qualifiers  
/organism="Homo sapiens (human)".  
1..697  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 52.3%; Score 567; DB 6; Length 697;  
Best Local Similarity 95.2%; Pred. No. 8,9e-120;

Matches	637;	Conservative	0;	Mismatches	24;	Indels	8;	Gaps	5;
Qy	14	GGGCGAGGAGGAGACAGCGCAAGGGGCGGAGATGGGCGGGCGGGCGGGCTCGG	73						
Db	19	GCGCAGGAGGATGACAGCGCAAGGGGCGGAGATGGGCGGGCGGGCGGGCTCGG	78						
Qy	74	GGAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	133						
Db	79	GGAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	138						
Qy	134	TCCCGGGCTTCGCGCATCCCTCCCGGCGCACTGGCGGGTCCCTCAAGCGCTTGAC	193						
Db	139	TCCCGGGCTTCGCGCATCCCTCCCGGCGCACTGGCGGGTCCCTCAAGCGCTTGAC	198						
Qy	194	TTCCGCTCAAAACCTGATCTCTTATGTCAGATGATGATGATGATGATGATGATG	253						
Db	199	TTCCGCTCAAAACCTGATCTCTTATGTCAGATGATGATGATGATGATGATGATG	258						
Qy	254	CCTATCCAGTATGAGGGTGTATGATGATGATGATGATGATGATGATGATGATG	313						
Db	259	CCTATCCAGTATGAGGGTGTATGATGATGATGATGATGATGATGATGATGATG	318						
Qy	314	TGGGATTTAATATGAGACCTCTGCGGACCTTGAAATTAATGATGATGATGATG	373						
Db	319	TGGGATTTAATATGAGACCTCTGCGGACCTTGAAATTAATGATGATGATGATG	378						
Qy	374	TTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	433						
Db	379	TTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	438						
Qy	434	GGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	493						
Db	439	GGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	498						
Qy	494	GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	552						
Db	499	GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	558						
Qy	553	AGTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	610						
Db	559	AGTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	618						
Qy	611	CAGG---ACATGAAACAGGATTTATATGAGA--CATGAAATGTAAGGCGGCGAG	665						
Db	619	CAGGAAATGTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG	678						
Qy	666	AAAAAGGGG 674							
Db	679	AAAAAANGG 687							

RESULT 12  
AC001226 106988 bp DNA linear PRI 29-MAY-1997  
AC001226  
AC001226.1 GI:2133862  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 106988)  
Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,  
Fasman, K.H. and Lander, E.S.  
Genomic sequence from Human 13

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
2 (bases 1 to 106988)  
Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,  
Fasman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,  
Barna, N., Brown, K., Cooke, P., Daly, M.J., Forrest, C., Frapp, M.J.,  
Gage, D., Geraghty, K., Hagos, B., Jaccotot, L., Lane, M., Mackenzie, J.,  
Marquis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A.,

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J., and  
Stilwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and  
Zody, M.  
Direct Submission  
Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 106988)  
Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,  
Fasman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,  
Barna, N., Chang, A., Cooke, P., Daly, M.J., Forrest, C., Frapp, M.J.,  
Gage, D., Geraghty, K., Hagos, B., Jaccotot, L., Lane, M., Mackenzie, J.,  
Marquis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A.,  
Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J., and  
Stilwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and  
Zody, M.

TITLE  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 29, 1997 this sequence version replaced gi:1932720.  
The Staden databases, finishing information, and all  
chromatographic files used in the assembly of this clone are  
available from our anonymous ftp site.

FEATURES  
source  
All repeats were identified using RepeatMasker: Smit, A.F.A. &  
Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.  
Location/Qualifiers  
1..106988  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="76N15"

repeat\_region  
complement(113..200)  
/rpt\_family="TH1B"  
repeat\_region  
complement(202..611)  
/rpt\_family="L1PA3"  
repeat\_region  
complement(674..694)  
/rpt\_family="AT\_rich"  
complement(835..981)  
repeat\_region  
complement(1632..1920)  
/rpt\_family="MIR2"  
repeat\_region  
complement(1970..2271)  
/rpt\_family="AluB"  
repeat\_region  
complement(2779..2946)  
/rpt\_family="AluSp"  
repeat\_region  
complement(3552..3828)  
/rpt\_family="FAM"  
repeat\_region  
complement(4361..4640)  
/rpt\_family="AluBp"  
repeat\_region  
complement(5253..5385)  
/rpt\_family="AluSg"  
repeat\_region  
complement(5914..6106)  
/rpt\_family="MLTIF"  
repeat\_region  
complement(6341..6435)  
/rpt\_family="MER34"  
repeat\_region  
complement(6838..7140)  
/rpt\_family="AluX"  
repeat\_region  
complement(7394..7455)  
/rpt\_family="MER5A"  
repeat\_region  
complement(7464..7627)  
/rpt\_family="MER5B"  
repeat\_region  
complement(7624..7792)  
/rpt\_family="MER5B"  
repeat\_region  
complement(8363..8391)  
/rpt\_family="AT\_rich"  
repeat\_region  
complement(8480..8639)  
/rpt\_family="MER5B"  
repeat\_region  
complement(9199..9286)  
/rpt\_family="TAN"  
repeat\_region  
complement(9340..9644)



```
/rpt_family="AluJo"
repeat_region 9880..9946
/rpt_family="MIR2"
repeat_region 10208..10787
/rpt_family="L1MB3"
repeat_region 10984..11297
/rpt_family="AluSg"
repeat_region 11311..11470
/rpt_family="L1PA8"
repeat_region 11471..11495
/rpt_family="AT_rich"
repeat_region 12394..12440
/rpt_family="AT_rich"
repeat_region 13782..13921
/rpt_family="MIR"
repeat_region 14035..14316
/rpt_family="AluSx"
repeat_region 15018..15046
/rpt_family="AT_rich"
repeat_region 15080..15373
/rpt_family="Aluub"
repeat_region 15388..15881
/rpt_family="AluJo"
repeat_region 18038..18071
/rpt_family="CAT)n"
repeat_region 20389..20434
/rpt_family="FAM"
repeat_region 21338..21534
/rpt_family="AluSx"
repeat_region 23331..23353
/rpt_family="GC_rich"
repeat_region 23546..23578
/rpt_family="GC_rich"
repeat_region 24565..24707
/rpt_family="MIR"
repeat_region 26013..26047
/rpt_family="MIR2"
repeat_region 27717..28532
/rpt_family="L1PB1"
repeat_region 28553..28854
/rpt_family="AluSg"
repeat_region 28860..29076
/rpt_family="L1"
repeat_region 29365..29561
/rpt_family="MIR"
repeat_region 30050..30147
/rpt_family="MIR"
repeat_region 31487..31812
/rpt_family="AluJo"
repeat_region 33629..33732
/rpt_family="MIR"
repeat_region 34434..34729
/rpt_family="AluSx"
repeat_region 35070..35361
/rpt_family="AluSx"
repeat_region 36437..36461
/rpt_family="AT_rich"
repeat_region 39967..40261
/rpt_family="AluSx"
repeat_region 40741..40915
/rpt_family="AluSg"
repeat_region 40963..41663
/rpt_family="L1PA2"
repeat_region 42042..42352
/rpt_family="MLT1E"
repeat_region 42467..42771
/rpt_family="AluSg"
repeat_region 42772..42848
/rpt_family="L1MB8"
repeat_region 42849..43151
/rpt_family="Aluub"
repeat_region 43152..43318
/rpt_family="L1MA10"
```

```
repeat_region 43253..43427
/rpt_family="L1MB8"
repeat_region 43429..43672
/rpt_family="L1MB8"
repeat_region 44188..44481
/rpt_family="AluSx"
repeat_region 44541..45246
/rpt_family="L1TR8"
repeat_region 46999..47032
/rpt_family="AT_rich"
repeat_region 47411..47431
/rpt_family="AT_rich"
repeat_region 47612..47684
/rpt_family="CAT)n"
repeat_region 47946..48239
/rpt_family="AluJo"
repeat_region 49827..49872
/rpt_family="AT_rich"
repeat_region 50160..50186
/rpt_family="AT_rich"
repeat_region 50952..51121
/rpt_family="MERSA"
repeat_region 52140..52284
/rpt_family="AluJo"
repeat_region 53772..53974
/rpt_family="MIR1B"
repeat_region 54225..54346
/rpt_family="MIR"
repeat_region 54533..54579
/rpt_family="TGAA)n"
repeat_region 54580..54736
/rpt_family="Aluub"
repeat_region 54755..54782
/rpt_family="CA)n"
```

```
Query Match 47.3%; Score 512.2; DB 9; Length 106988;
Best Local Similarity 99.4%; Pred. No. 1.3e-106;
Matches 514; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 568 TATATCAGAAACATGTTCAACCAATGGCAAGGTTGAAACAGACATGAAACAG 627
Db 31859 TAAACTAGGAAACATGTTCAACCAATGGCAAGGTTGAAACAGACATGAAACAG 31918

QY 628 AATTTATATGACATGCAATGTAAGCCAGCCAGAAAGGGGCGAGACATGTT 687
Db 31919 AATTTATATGACATGCAATGTAAGCCAGCCAGAAAGGGGCGAGACATGTT 31978

QY 688 TGATTCCTAGCACTGTTCCAAATTTGTTAAGACCTTTAACAAGTGGCTGAATTGG 747
Db 31979 TGATTCCTAGCACTGTTCCAAATTTGTTAAGACCTTTAACAAGTGGCTGAATTGG 32038

QY 748 AGCAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTTACATGGAGAAC 807
Db 32039 AGCAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTTACATGGAGAAC 32098

QY 808 TACTTATCTGGGAAATGAAACATCTGTTTGGGCCCAACAGGAAACAAGCTCTGGTT 867
Db 32099 TACTTATCTGGGAAATGAAACATCTGTTTGGGCCCAACAGGAAACAAGCTCTGGTT 32158

QY 868 AGCATAAAGAAATTTATATACCCCTTCAACACCAATTTGCCAATTAAGATTTGCTT 927
Db 32159 AGCATAAAGAAATTTATATACCCCTTCAACACCAATTTGCCAATTAAGATTTGCTT 32218

QY 928 GAGTCTCTGCAAAATTTTGATGCAAGTATGTCACAAACAGTCTATTTGTTATAA 987
Db 32219 GAGTCTCTGCAAAATTTTGATGCAAGTATGTCACAAACAGTCTATTTGTTATAA 32278

QY 988 TTTTGAATATGGTTTATACCTATGAAATTTCCCTTTTATTAATAACATAGAAAT 1047
Db 32279 TTTTGAATATGGTTTATACCTATGAAATTTCCCTTTTATTAATAACATAGAAAT 32338

QY 1048 CCTTACCTATCAGAAACAAACACTCTGTTTA 1084
```

Db 32339 CCCTTACTATCAGAAACAAACACTCTCTGTTTA 32375

RESULT 13  
AL136440 169362 bp DNA linear HTG 10-JUL-2001  
LOCUS Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.  
DEFINITION AL136440  
ACCESSION AL136440.2 GI:10039473  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS Burton, J.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
requests: clonerequest@sanger.ac.uk  
On Sep 8, 2000 this sequence version replaced gi:6982057.

COMMENT  
----- Genome Center  
Center: Sanger Centre  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA185124  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; 108752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 16208 bases at least Q40  
Consensus quality: 164521 bases at least Q30  
Consensus quality: 16040 bases at least Q20  
Insert size: 168062; sum-of-contigs  
Insert size: 164996; 1.7% error; agarose-fp  
Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality  
coverage: 4.16x in Q20 bases; agarose-fp

----- NOTE: This is a 'working draft' sequence. It currently  
\* consists of 14 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 3400: contig of 3400 bp in length  
\* 3401 3500: gap of 100 bp  
\* 16445 16444: contig of 12944 bp in length  
\* 16445 16444: gap of 100 bp  
\* 16445 16444: contig of 2970 bp in length  
\* 19515 19514: gap of 100 bp  
\* 19515 19514: contig of 2627 bp in length  
\* 22241 22241: gap of 100 bp  
\* 22242 22241: contig of 3185 bp in length  
\* 25526 25526: gap of 100 bp  
\* 25527 25526: gap of 100 bp  
\* 25527 25526: contig of 7444 bp in length  
\* 33071 33070: gap of 100 bp  
\* 33071 33070: contig of 34450 bp in length  
\* 67621 67620: gap of 100 bp  
\* 67621 67620: contig of 25693 bp in length  
\* 93414 93413: gap of 100 bp  
\* 93414 93413: contig of 20368 bp in length  
\* 114081 114081: gap of 100 bp  
\* 114081 114081: contig of 12025 bp in length  
\* 114182 114182: gap of 100 bp  
\* 126307 126306: gap of 100 bp  
\* 126307 126306: contig of 18277 bp in length  
\* 144584 144583: gap of 100 bp  
\* 144584 144583: contig of 2453 bp in length  
\* 147137 147136: gap of 100 bp  
\* 147137 147136: contig of 17689 bp in length  
\* 147237 147236: gap of 100 bp  
\* 147237 147236: contig of 17689 bp in length

FEATURES  
\* 164926 165025: gap of 100 bp  
\* 165026 169362: contig of 4337 bp in length.

SOURCE  
1. 169362  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-185124"  
/clone\_1b="RP11-11.1"  
1. 3400  
/note="assembly\_fragment:00318  
clone end:77  
vector side:left"

misc\_feature  
3501..16444  
/note="assembly\_fragment:00726  
fragment\_chain:1"

misc\_feature  
16545..19514  
/note="assembly\_fragment:00157  
fragment\_chain:1"

misc\_feature  
19615..22241  
/note="assembly\_fragment:00846  
fragment\_chain:1"

misc\_feature  
22342..25526  
/note="assembly\_fragment:00271  
fragment\_chain:1"

misc\_feature  
25627..33070  
/note="assembly\_fragment:01555  
fragment\_chain:1"

misc\_feature  
33171..67620  
/note="assembly\_fragment:00768"  
67721..93413  
/note="assembly\_fragment:00828"  
93514..114081  
/note="assembly\_fragment:01026"  
114182..126206  
/note="assembly\_fragment:01600  
fragment\_chain:2"

misc\_feature  
126307..144583  
/note="assembly\_fragment:01198  
fragment\_chain:2"

misc\_feature  
144684..147136  
/note="assembly\_fragment:01132  
fragment\_chain:2"

misc\_feature  
147237..164925  
/note="assembly\_fragment:01005  
fragment\_chain:2"

misc\_feature  
165026..169362  
/note="assembly\_fragment:01413  
clone end:SP6  
vector\_side:right"

ORIGIN  
Query Match 47.3% Score 512.2; DB 2; Length 169362;  
Best Local Similarity 99.4%; Pred. No. 1.5e-106;  
Matches 514; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

568 TATATCAGGAAACATGTTCAACCAATGCGCAAGTGGTGAACACGCAATGAACACG 627  
Db 118303 TAAATCAGGAAACATGTTCAACCAATGCGCAAGTGGTGAACACGCAATGAACACG 118362

628 AATTATATGAGACATGAAATGTAAAGCCAGCCAGAAAAGGGGAGAGACATGTT 687  
Db 118363 AATTATATGAGACATGAAATGTAAAGCCAGCCAGAAAAGGGGAGAGACATGTT 118422

688 TGATTCTCAGACGCTGTTCCAAATTTGTGTAAAGACCTTTAACAAGTTGCTGAATTTGG 747  
Db 118423 TGATTCTCAGACGCTGTTCCAAATTTGTGTAAAGACCTTTAACAAGTTGCTGAATTTGG 118482

748 AGCAGAGTTCAAGAACATGAACCAATATCAGAAATATTTCTTTACAGTGAAGAAC 807  
Db 118483 AGCAGAGTTCAAGAACATGAACCAATATCAGAAATATTTCTTTACAGTGAAGAAC 118542



```

* 187239 187338: gap of 100 bp
* 187339 189602: config of 2266 bp in length
* 189603 189702: gap of 100 bp
* 189703 194463: config of 4761 bp in length
* 194464 194563: gap of 100 bp
* 194564 197682: config of 3119 bp in length
* 197683 197782: gap of 100 bp
* 197783 201788: config of 4006 bp in length
* 201789 204060: gap of 100 bp
* 204061 204160: config of 2172 bp in length
* 204161 206237: gap of 100 bp
* 206238 210376: config of 2077 bp in length
* 210377 210476: gap of 100 bp
* 210477 225245: config of 14769 bp in length
* 225246 230027: gap of 100 bp
* 230028 230127: config of 4682 bp in length
* 230128 23610: gap of 100 bp
* 23611 232710: config of 2483 bp in length
* 232711 234942: gap of 100 bp
* 234943 235042: config of 2232 bp in length
* 235043 240958: gap of 100 bp
* 240959 241058: config of 5916 bp in length
* 241059 251187: gap of 100 bp

```

```

FEATURES
    source
        1. 251187
            location/qualifiers
                1. 251187
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="13"
                    /clone_id="RP11-453N22"
                    /clone_id="RP11-453N22"

```

```

misc_feature
    /note="assembly fragment:00469"
    fragment_chain:1"
    2231..5689
misc_feature
    /note="assembly fragment:01219"
    fragment_chain:1"
    5790..10046
misc_feature
    /note="assembly fragment:00352"
    fragment_chain:1"
    10147..13624
misc_feature
    /note="assembly fragment:00547"
    fragment_chain:2"
    12725..20993
misc_feature
    /note="assembly fragment:00617"
    fragment_chain:2"
    21094..28936
misc_feature
    /note="assembly fragment:01248"
    fragment_chain:2"
    29037..33230
misc_feature
    /note="assembly fragment:00606"
    fragment_chain:3"
    33231..35153
misc_feature
    /note="assembly fragment:01016"
    fragment_chain:3"
    35254..37649
misc_feature
    /note="assembly fragment:01277"
    fragment_chain:3"
    37750..40245
misc_feature
    /note="assembly fragment:00920"
    fragment_chain:4"
    40346..48420
misc_feature
    /note="assembly fragment:01288"
    fragment_chain:4"
    48521..52169
misc_feature
    /note="assembly fragment:00615"
    fragment_chain:4"
    52270..56443
misc_feature
    /note="assembly fragment:00009"
    fragment_chain:5"

```

```

misc_feature 56544..63088
    /note="assembly fragment:00448"
    fragment_chain:5"
    63189..65281
    /note="assembly fragment:00027"

Query Match
Best local similarity 99.4%, Pred. No. 1,7e-106;
Matches 514; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

misc_feature
    /note="assembly fragment:00027"

568 TATATCAGAAACATGTTCAACCAATGCGAAGTGGTGAACACAGACATGAACACAG 627
Db TAACTAGAAACATGTTCAACCAATGCGAAGTGGTGAACACAGACATGAACACAG 194040

628 AATTAATATAGACATGATGATGAAGACACCCAGAAAGGGGCGAGACATGTT 687
Db AATTAATATAGACATGATGATGAAGACACCCAGAAAGGGGCGAGACATGTT 193980

688 TGATTCCTACACATGTTCCAAATTTGTTTACGACCTTTACAGATGGCTGAATTTGG 747
Db TGATTCCTACACATGTTCCAAATTTGTTTACGACCTTTACAGATGGCTGAATTTGG 193920

748 AGCAGAGTTCAAGAACATGAACCAACCTATACAAATATTTCTTACAGTGAACAC 807
Db AGCAGAGTTCAAGAACATGAACCAACCTATACAAATATTTCTTACAGTGAACAC 193860

808 TACTATCTGGGAATGAACATCTGTTTGGGCGCAACAGAAACAAACATCTTGTT 867
Db TACTATCTGGGAATGAACATCTGTTTGGGCGCAACAGAAACAAACATCTTGTT 193800

868 AGCCATTAAGATTTATTAACCCCTCAACACATTTGCCACTMAAGATTTCTGTT 927
Db AGCCATTAAGATTTATTAACCCCTCAACACATTTGCCACTMAAGATTTCTGTT 193740

928 GAGTCTCTGCAATTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 987
Db GAGTCTCTGCAATTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 193680

988 TTTTGAATATGTTGTTTATCTATGAATATCCCTTTATTAATTAATTAATTAAT 1047
Db TTTTGAATATGTTGTTTATCTATGAATATCCCTTTATTAATTAATTAATTAAT 193620

1048 CCCTTATCTATCAAGAAACCAACACCTCTGTTTA 1084
Db CCCTTATCTATCAAGAAACCAACACCTCTGTTTA 193563

193619 CCCTTATCTATCAAGAAACCAACACCTCTGTTTA 193563

RESULT 15
BD060465 506 bp DNA linear PAT 27-AUG-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 506)
REFERENCE
AUTHORS
Jacobs, K., McCoy, J.M., Lavallie, E.R., Racie, L.A., Metberg, D.,
Treacy, M., Spaulding, V., and Agostino, M.J.
Secreted expressed sequence tags (seSTs).
Patent: JP 2001518793-A 825 16-OCT-2001;
GENETICS INSTITUTE INC
PN JP 2001518793-A/825
PD 16-OCT-2001
PF 10-APR-1998 JP 1998543070
PR 10-APR-1997 US 08/837312
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
DAVID MERRBERG
PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
C12N15/12, C12N5/10, C07K14/47, C12Q1/68, A61K38/17 CC Strandness:

```

Double;  
CC Topology: Linear;  
FH key Location/Qualifiers.  
FEATURES  
source Location/Qualifiers  
1..506  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4577"

## ORIGIN

Query Match 37.0%; Score 401.4; DB 6; Length 506;  
Best Local Similarity 99.8%; Pred. No. 1.1e-81;  
Matches 402; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 497 GCTGCTGCTTTTGGAGGATTTGATGATGTTCTCACTGGAGAAAATGGACATTAGTT 556
    |||||||
Db 99 GCTGCTGCTTTTGGAGGATTTGATGATGTTCTCACTGGAGAAAATGGACATTAGTT 158

QY 557 CAAGTAGCACTATATCAGAAAACATGTTCAACCAATGGCAAAAGTGGTGAACAGGAC 616
    |||||||
Db 159 CAAGTAGCACTATATCAGAAAACATGTTCAACCAATGGCAAAAGTGGTGAACAGGAC 218

QY 617 AATGAAACAGAAATTTATATGAGACATGAAATGTAAAAAGCCAGCCCAAGAAAAGGGGCA 676
    |||||||
Db 219 AATGAAACAGAAATTTATATGAGACATGAAATGTAAAAAGCCAGCCCAAGAAAAGGGGCA 278

QY 677 GAGACATGTTGATTTCCATGACACTGTTCCAAATTTGTTAAGGACCTTTAACAAGTTG 736
    |||||||
Db 279 GAGACATGTTGATTTCCATGACACTGTTCCAAATTTGTTAAGGACCTTTAACAAGTTG 338

QY 737 GCTGAATTTGAGCAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTTAC 796
    |||||||
Db 339 GCTGAATTTGAGCAGAGTTCAAGAACATAGAAACCAACTATACAGAAATATTTCTTAC 398

QY 797 AGTGAAGAACTTACTATCTGGAAAATGAAACATCTGTTTTTGGGCCAACAGAAAACAAG 856
    |||||||
Db 399 AGTGAAGAACTTACTATCTGGAAAATGAAACATCTGTTTTTGGGCCAACAGAAAACAAG 458

QY 857 ACTCTGGTTTAGCCATAAAAAGATTTTATTTACCCCTTCAAG 899
    |||||||
Db 459 ACTCTGGTTTAGCCATAAAAAGATTTTATTTACCCCTTCAAGC 501
```

Search completed: April 22, 2004, 02:57:20  
Job time : 4425.69 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

3M nucleic - nucleic search, using sw model

Run on: April 22, 2004, 00:17:22 ; Search time 2854.26 Seconds

(Without alignments)  
9918.266 Million cell updates/sec

Title: US-10-010-050A-1\_COPY\_137\_1084

Perfect score: 948

Sequence: 1 cgggctccgggcatccctc.....acaaacactctgttta 948

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_ests1:\*  
2: em\_ests2:\*  
3: em\_ests3:\*  
4: em\_ests4:\*  
5: em\_ests5:\*  
6: em\_ests6:\*  
7: em\_ests7:\*  
8: em\_ests8:\*  
9: em\_ests9:\*  
10: em\_ests10:\*  
11: em\_ests11:\*  
12: em\_ests12:\*  
13: em\_ests13:\*  
14: em\_ests14:\*  
15: em\_ests15:\*  
16: em\_ests16:\*  
17: em\_ests17:\*  
18: em\_ests18:\*  
19: em\_ests19:\*  
20: em\_ests20:\*  
21: em\_ests21:\*  
22: em\_ests22:\*  
23: em\_ests23:\*  
24: em\_ests24:\*  
25: em\_ests25:\*  
26: em\_ests26:\*  
27: em\_ests27:\*  
28: em\_ests28:\*  
29: em\_ests29:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	864.8	91.2	1201 9	AL546472 AL546472
2	691.8	72.0	13	BX31615 BX31615
3	683.6	73.1	13	BX34262 BX34262
4	674.2	71.1	1201 9	AL571805 AL571805

Result No.	Score	Query Match Length	ID	Description
5	673	71.0	931	BE873363
6	671.8	70.9	2356	AK043247
7	670.6	70.7	689	BU621797
8	670.2	70.7	2433	AK085741
9	668.6	70.5	2247	AK032293
10	645.4	68.1	908	BX370641
11	643.2	67.8	924	CA488543
12	629.6	66.4	869	BI090566
13	629.6	66.4	869	BI090566
14	601	63.4	787	BI223533
15	600	63.3	787	BI223533
16	562.4	59.3	1098	BF982158
17	560.4	58.1	915	BI080303
18	546.2	57.6	554	CB297292
19	542.6	57.2	755	BC443453
20	522.8	55.1	974	BO319385
21	507.8	53.6	676	BI223533
22	506.2	53.4	938	BO326170
23	489.4	51.6	607	AM475730
24	475	50.1	607	BM849262
25	474.6	50.1	828	BI156421
26	468.8	49.5	872	BI734549
27	468.2	49.4	833	CK128221
28	458.4	48.4	796	BI182240
29	455.6	48.1	853	BI456284
30	447.2	47.2	623	BO417448
31	442.4	46.7	599	AM594284
32	435.6	45.9	706	BF739857
33	435	45.9	516	BF739857
34	430	45.4	633	BM740661
35	429	45.3	565	AM280382
36	423.8	44.7	690	CE724176
37	421.2	44.4	484	AV723592
38	399.6	42.2	854	BI082103
39	395	41.7	791	BG619558
40	385.8	40.7	661	BY734042
41	383.4	40.4	465	BI304716
42	382	40.3	825	BU633296
43	380.6	40.1	767	AT451053
44	379.2	40.0	915	AT451053
45	377.6	39.8	3269	AK030852

## ALIGNMENTS

RESULT 1  
AL546472  
LOCUS  
DEFINITION  
AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
clone CS0D1030Y01 5-PRIME, mRNA sequence.  
ACCESSION  
AL546472 GI:31268306  
VERSION  
AL546472.2  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
COMMENT  
On Feb 15, 2001 this sequence version replaced gi:12679620.  
Contact: Genoscope - Centre National de Sequencage  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1030Y01&cluster=7238.f. Contact :  
Feng Liang Email : liang@lifetech.com URL :



http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D1030CE01Qp1.  
Location/Qualifiers  
1. 1201

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1030Y01"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_id="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 91.2%; Score 864.8; DB 9; Length 1201;  
Best Local Similarity 95.4%; Pred. No. 6.5e-186;  
Matches 904; Conservative 17; Mismatches 19; Indels 8; Gaps 3;

```
OY 1 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 60
DB 153 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 212
OY 61 CGTCCAAAACCTGATCTTATTTGTCAGAGTAAATATCTTCTGTCGAAGTGGCTCACT 120
DB 213 CGTCCAAAACCTGATCTTATTTGTCAGAGTAAATATCTTCTGTCGAAGTGGCTCACT 272
OY 121 ATCCCAAGTTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 273 ATCCCAAGTTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 332
OY 181 GAATTTAAATATGAGACCTCTCGGGAACCTGGAATTAATGATGATGATGATGATGATG 240
DB 333 GAATTTAAATATGAGACCTCTCGGGAACCTGGAATTAATGATGATGATGATGATGATG 392
OY 241 AGAAGTACATTAAGTGGAGAACTGACATGATGATGATGATGATGATGATGATGATGATG 300
DB 393 AGAAGTACATTAAGTGGAGAACTGACATGATGATGATGATGATGATGATGATGATGATG 452
OY 301 AACTGTACATTTCCCACTCTCGGACCTGGAATGATGATGATGATGATGATGATGATGATG 350
DB 453 AACTGTACATTTCCCACTCTCGGACCTGGAATGATGATGATGATGATGATGATGATGATG 512
OY 361 GGTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 513 GGTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 572
OY 421 CAAGTACGATATATCAAGAAACATGTTCAACCAATGGCAAGTGGTGAACAGAGAC 430
DB 573 CAAGTACGATATATCAAGAAACATGTTCAACCAATGGCAAGTGGTGAACAGAGAC 632
OY 481 AATGAAACAGAAATTTATATGAGCATGATGATGATGATGATGATGATGATGATGATG 540
DB 633 AATGAAACAGAAATTTATATGAGCATGATGATGATGATGATGATGATGATGATGATG 692
OY 541 GAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
DB 693 GAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 752
OY 601 GCTGAATTTGAGAGAGTTCAGAAACATGAAACCAATATCAAGATATTTCTTAC 650
DB 753 GCTGAATTTGAGAGAGTTCAGAAACATGAAACCAATATCAAGATATTTCTTAC 812
OY 661 AGTGAAGAACTGATATCTGAGGAATGAAACATGATGATGATGATGATGATGATGATG 720
DB 813 AGTGAAGAACTGATATCTGAGGAATGAAACATGATGATGATGATGATGATGATGATG 872
OY 721 ACTCTGTTGTTAGCAATAAAGATTTTATTTACCCCTCAACCACTTTGGCACTTAA 730
DB 873 ACTCTGTTGTTAGCAATAAAGATTTTATTTACCCCTCAACCACTTTGGCACTTAA 932
OY 781 GAATTTCTGTTGATGCTCTGCAAAATTTTGTATGATGATGATGATGATGATGATGATG 840
```

```
DB 933 GAATTTCTGTTGATGCTCTGCAAAATTTTGTATGATGATGATGATGATGATGATGATG 992
OY 841 TTGTTTATATTTTGAATATGTTTGTATGATGATGATGATGATGATGATGATGATGATG 900
DB 993 TTGTTTATATTTTGA--TATGCTTTTTCATGATGATGATGATGATGATGATGATGATG 1047
```

```
OY 901 TATGAGAAATCCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 948
DB 1048 TATGAGAAATCCCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
```

## RESULT 2

EX331615 923 bp mRNA linear EST 01-MAY-2003  
DEFINITION BX331615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens  
CDNA clone CS0D1009Y01 5-PRIME, mRNA sequence.  
ACCESSION BX331615  
VERSION BX331615.1 GI:30310073  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 923)  
AUTHORS Li, W.B., Gruber, C., Jessup, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1009AB01Qp1&cluster=7238.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D1009AB01Qp1.

## FEATURES

## source

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1009Y01"  
/tissue\_type="NEUROBLASTOMA COT 10-NORMALIZED"  
/clone\_id="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 73.0%; Score 691.8; DB 13; Length 923;  
Best Local Similarity 96.5%; Pred. No. 1.2e-146;  
Matches 690; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

```
OY 1 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 60
DB 208 CGGGTCTCGGGGATCCCTCCCGGGCCCACTGGCCGGCCCTTACAGAGGCTTGACTTC 267
OY 61 CGTCCAAAACCTGATCTTATTTGTCAGAGTAAATATCTTCTGTCGAAGTGGCTCACT 120
DB 268 CGTCCAAAACCTGATCTTATTTGTCAGAGTAAATATCTTCTGTCGAAGTGGCTCACT 327
OY 121 ATCCCAAGTTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
DB 328 ATCCCAAGTTATGAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATG 387
OY 181 GAATTTAAATATGAGACCTCTCGGGAACCTGGAATTAATGATGATGATGATGATGATG 240
DB 388 GAATTTAAATATGAGACCTCTCGGGAACCTGGAATTAATGATGATGATGATGATGATG 447
```

	ly	241	AACAATCATTTAACTGCGCAAACTACACAAATGGAAAGGAATGAACCTTTCACAACTGGC	300
	lb	448	AGAATTATTTAATCTGGGAAGAACTTAACAATGGAATGGATTAACCTTTCCACTTGGG	507
	ly	301	AACTGTAACTTCCCACATCTCCAGCTGAAAATGAGATGCCCTTTTGATGTAAATCAAGGC	360
	lb	508	AANTGTANAATTTCCCGCTCTCCGANNCTGNAATGATGSSCTTTTCTGGKGTATCAAGGG	567
	ly	361	GCTGCTCGCTTTTTTAGAGGAATGTANGATGTTTCACTGAAAGGAAAAATGGACATTAAGTT	420
	lb	568	GSATGCTGTTTTTGTGRGGAAATTTGRTGAAGTTTCACTGAAAGGAAAAATGGACATTAAGTT	627
	ly	421	CAGATGACAACTATATCATAGAAAACATGTTCAACCMAATGGCAAAGTGGGTGAAACAGAC	480
	lb	628	CAAGTAGCACTATATCATGAGAAAACATGTTCAACCMAATGGCAAAGTGGGTGAAACAGAC	687
	ly	481	AATGAAACGAGAAATTTATTAAGACATGAAATGTAAAGCCAGCCCAAGAAAAGGGGCGCA	540
	lb	688	AATGAAACGAGAAATTTATTAAGACATGAAATGTAAAGCCAGCCCAAGAAAAGGGGCGCA	747
	ly	541	GAGACATGTTTGAATTCCTACGACTGTCCAAATTTGTGTAAAGACCTTTAACAAAGTTG	600
	lb	748	GAGACATGTTTGAATTCCTACGACTGTCCAAATTTGTGTAAAGACCTTTAACAAAGTTG	807
	ly	601	GCTGTAATTTGAGCAGAGTTCAAGAACATAGAAAACCAATATCAAGAAATATTTCTTTAC	660
	lb	808	GCTGTAATTTGAGCAGAGTTCAAGAACATAGAAAACCAATATCAAGAAATATTTCTTTAC	867
	ly	661	AGTGGAGAACTTACTTATCTGGGAATGAAACATCTGTCTTTTGGGCCAACAGGAA	715
	lb	868	AGTGGAGAACTTACTTATCTGGGAATGAAACATCTGTCTTTTGGGCCAACAGGAA	922

RESULT 3  
BX42662  
ACUS  
DEFINITION BX42662 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
ACCESSION BX42662  
VERSION BX42662.1 GI:30338103  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eubacteria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1077)  
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f For  
more information about this cluster, see  
[http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODL004CH02QPI&cluster=7238.f](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODL004CH02QPI&cluster=7238.f). Contact :  
Feng Liang Email : [Eliang@lifetech.com](mailto:Eliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paraday Avenue Genoscope sequence ID : CSODL004CH02QPI.  
Location/Qualifiers  
1..1077  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODL004YP03"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_1tb="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
25-NORMALIZED"  
/note="1st strand cDNA was primed with a Notti-oligo(dT)

FEATURES  
source

ORIGIN

primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and BcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match	72.1%;	Score 683.6;	DB 13;	Length 1077;
Best Local Similarity	88.8%;	Pred. No. 8.8e-145;		
Matches 700; Conservative	49;	Mismatches 34;	Indels 5;	Gaps 2;

QY		1	CGGGCTCCGGGCAMC	CCCCCTCCCGGCGCA	CTGGCCGGTGCCCTACAGACGCTTTGACTTC	60
Db		206	CGGGTCTCGGGCAT	TCCCCTCCCGGCGCA	CTGGCCGGTGCCTACAGCGCTTTGACTTC	265
QY		61	CGTCCAAACCCTAT	CCTTATTGTCAAGCTA	GATATCTTTCTGTGCCACTGCTCACT	120
Db		266	CGTCCAAAACCTAT	CTCTTATTGTCAAGCTA	GATATCTTTCTGTGCCACTGCTCACT	325
QY		121	ATCCCACTTA	TGAGGGGTGATGATGACATTGAAGTTTTTGATTACAAGCCCACTGATGG	180	
Db		326	ATCCCACTTATGAGA	GGGTGATGATGACATTGAAGTTTTTGATTACAAGCCCACTGATGG	385	
QY		181	GAATTTTAAATATGAG	CCCTCCGGGCACTTGAAATTTNGCATGATGOCATGTGATTC	240	
Db		386	GAATTTTAAATATGAG	CCCTCCGGGCACTTGAAAAATTTNGCATGAGGCCATGTGATTC	445	
QY		241	AGAAGTACATTA	CTGSCAAGAACTACA	CAATGGAATGATGATGACTTTTCCAATCTGGGC	300
Db		446	AGAAGTACATTA	CTGSCAAGAACTACA	CAATGGAATGATGATGACTTTTCCAATCTGGGC	505
QY		301	AACGTACATTTT	CCCCACTCCGACCTGAAT	TGATGCCCTTTCTGTGTATCAAAGC	360
Db		506	AACGTACATTTT	CCCCACTCCGACCTGAAT	TGATGCCCTTTCTGTGTATCAAAGC	565
QY		361	GCCTGCTGCTTTT	TGAGGGAATGATGATGTTCACTGGAAGGAAATGGGACATTAGTT	420	
Db		566	GCCTGCTGCTTTT	TGAGGGAATGATGATGTTCACTGGAAGGAAATGGGACATTAGTT	625	
QY		421	CAAGTAGCAAT	CTATPACGAGAACAT	GTGTCAACCAATGGCAAGTGGGTGAAACAGGAC	480
Db		626	CAAGTAGCAAT	CTATPACGAGAACAT	GTGTCAACCAATGGCAAGTGGGTGAAACAGGAC	685
QY		481	AATGAAACAGGA	TTTATTTATGAGACATGGAATGATAAACCCAGCCGACGAAAGGGGGCA	540	
Db		686	AATGAAACAGRA	TTTATTTATGAGACATKKAATGTMAAAACARCCCAAAAAGGGGSA	745	
QY		541	GAGACATG	GTGTAATCTTACGACTGTCCAAATTTTGTTGAAGACCTTTAACAAATGG	600	
Db		746	GAGACATG	GTGTAATCTTACGACTGTCCAAATTTTGTTGAAGACCTTTAACAAATGG	805	
QY		601	GCTGAATTTGAG	CAGAGTTCAAGAACTATGAAACCACTATACAGATATTTCTTTAC	660	
Db		806	GCTGAATTTTGA	CAGAGTTCAARAAAMAAAAACAAATATACAAAATATTTCTTTMY	865	
QY		661	AGTGGAGAACCTA	CTTATCTGGGAAATGAACATCTGTTTTTGGGCGCAACGAGAAACAG	720	
Db		866	AGTGGARAR-CTA	CTTATCTKGGAATGAAACATCTGTTTTTGGGSGACARGAACMAAA	924	
QY		721	ACTCTTGTTTAA	GCATAAAAAGATTTTATTAATCACCTTCAACACCATTTGGCCACTTAA	780	
Db		925	CTCTTGTTTAA	GCATAAAAAGATTTTWT-----SCCTCNARCCCMWTTGSCANHAAAAA	980	
QY		781	GAATTTCT	788		
Db		981	ATKXTT	988		
<hr/>						
RESULT 4						
AL571805/c		1201 bp	mRNA	linear	EST 31-MAY-2003	
LOCUS		AL571805 Homo sapiens	PLACENTA COT 25-NORMALIZED	Homo sapiens	CDNA	
DEFINITION		cGene CDSDI030YU01 3-PRIME, mRNA sequence.				
ACCESSION		AL571805				
VERSION		AL571805.2	GI:31293196			

KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Bokaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 1201)
TITLE	Li, M. B., Gruber, C., Jesse, J. and Polyes, D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	Unpublished (2001)
	On Feb 16, 2001 this sequence version replaced gi:12929467.

Genoscope - Centre National de Séquençage  
BP 191 91006 EVRY cedex - France  
Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f For  
more information about this cluster, see  
[http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f). Contact :  
Feng Liang Email: [fliang@lifetech.com](mailto:fliang@lifetech.com) URL: <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0D1030CE01NP1.  
Location/Qualifiers

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS010303701"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_1ib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/clone="1st strand cDNA was primed with a NotI-oligo (dt) primer. Five primers end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMSPORT 6 vector. Library was normalized.
```

## ORIGIN

Query Match	71.1%;	Score 674.2;	DB 9;	Length 1201;
Best Local Similarity	97.8%;	Pred. No. 1.2e-142;		
Matches 704; Conservative	9;	Mismatches 4;	Indels 3;	Gaps 3;

QY	230	CCATTGGATTCAAGAGTACATTAACCTGGCAAGAACTACACATGAGTAATGATGAACCTTT	289
	:	:	:
Db	1075	SCCATKGAATTCAAGAGTACATTAACCTGGCAA-RAMTACAAATAAGAAATGGATATG-AMTTT	1018
QY	290	TCCAACTTGGCAACATGTCATTTTCCCCCATCTCCGACCTGCAAAATGAATGCCCCCTTCTGGT	349
Db	1017	TCCAACTTGGCAACATGTCATTTTCCCCCATCTCCGACCTGCAAAATGAATGCCCCCTTCTGGT	958
QY	350	GTAAATCAAGGGCGTGCCTGCTTTTGTAGGG-AAATGATGATGTTCACTGGAAAGAAAT	408
Db	957	GTAAATCAAGGGCGTGCCTGCTTTTGTAGGGAAATGATGATGTTCACTGGAAAGAAAT	858
QY	409	GGGACATTAGTTCAAAGTAGCAACTATATCCAGAAACATGTTCAACCAATGGCAAAAGTG	468
Db	897	GGGACATTAGTTCAAAGTAGCAACTATATCCAGAAACATGTTCAACCAATGGCAAAAGTG	838
QY	469	GTGAAACAGGACCAATGAACAGGAATTTATATGACATGAGATGTAAACCCAGCCCA	528
Db	837	GTGAAACAGGACCAATGAACAGGAATTTATATGACATGAGATGTAAACCCAGCCCA	778
QY	529	GAAAAAGGGGGGACAGACATGAGTTGATTCCTGAAGACGTGCCAAATTTGTGTTAGAGCC	568
Db	777	GAAAAAGGGGGGACAGACATGAGTTGATTCCTGAAGACGTGCCAAATTTGTGTTAGAGCC	718
QY	589	TTTAAACAAGTTGCGCTGAATTTGGAGCAAGTTCAGAAACATAGAAACCAACTATACAGA	648
Db	717	TTTAAACAAGTTGCGCTGAATTTGGAGCAAGTTCAGAAACATAGAAACCAACTATACAGA	658
QY	649	AATATTTCTTTACAGTGGAGAACCTACTATCTGGGAAATGAACATCTGTTTTGGGCA	708
Db	657	AATATTTCTTTACAGTGGAGAACCTACTATCTGGGAAATGAACATCTGTTTTGGGCA	558
QY	709	ACAGGAAACAAGACTCTTGTTAGCCATAAAAAGATTTTATTAACCTTCAAAACCAAT	768

Db	597	ACAGAAACAGACCTCTGGTTTAGCCATAAAAAAGATTATTAATCCCTCTCAAAACACAT	538
Qy	769	TTGCCACTAAAGAAATTCCTGTAGCTCTCTGCATAATTTTGATGCAGTGAATTGCGAC	828
Db	537	TTGCCACTAAAGAAATTCCTGTAGCTCTCTGCATAATTTTGATGCAGTGAATTGCGAC	478
Qy	829	AAAACAGTTCTATTTGTTTATTAATTTGAAATATGTTTTTACCTATGAAATTCCTCTTT	888
Db	477	AAAACAGTTCTATTTGTTTATTAATTTGAAATATGTTTTTACCTATGAAATTCCTCTTT	418
Qy	889	ATTAAATTAACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGGTTTA	948
Db	417	ATTAAATTAACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGGTTTA	358

RESULT 5	BE873363	931 bp	mrna	linear	EST 20-OCT-2000
LOCUS	BE873363				
DEFINITION	BE873363	601450425F1 NIH MGC_65 Homo sapiens cDNA clone IMAGE:3854119 5',			
		mrna sequence.			
ACCESSION	BE873363				
VERSION	BE873363.1	GI:10322139			
KEYWORDS	EST.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				

REFERENCE	1 (bases 1 to 931)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

CDNA library Preparation: Life Technologies, Inc.  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LILNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LILNL at:  
<http://image.llnl.gov>  
plate: LLAM9579 row: b column: 08  
High quality sequence stop: 662.

**FEATURES**

**Source**

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3854119"
/risuse_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_1bp="NH_MGC_65"
/note="ORGAN: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 Kb. Library constructed by Life
Technologies."

```

## ORIGIN

Query Match	71.0%	Score 673	DB 10	Length 931
Best Local Similarity	93.0%	Pred. No. 2.3e-142		
Matches 727	Conservative 0	Mismatches 51	Indels 4	Gaps 2
QY	138	TGATGATGACATTGAGTTTTCGATATACAGCCCGAGATGGGAAATTAAATATGAGAGA	197	
Db	1	TGATGATGACATTGAGTTTTCGATATACAGCCCGAGATGGGAAATTAAATATGAGAGA	60	
QY	198	CTCTCTGGGACACTTGAAAATTATGATGATGCCATTGGATTGAGAGTACATTAACTGG	257	
Db	61	CTCTCTGGGACACTTGAAAATTATGATGATGCCATTGGATTGAGAGTACATTAACTGG	120	
QY	258	CAGGAAGCTACAAATGGAAATGGATATGAACTTTTCGAACCTTGGCAACGTACATTTCCCA	317	
Db	121	CAGGAAGCTACAAATGGAAATGGATATGAACTTTTCGAACCTTGGCAACGTACATTTCCCA	180	

```

2y 318 TCTCGAAGCTGAATGATGCGCCCTTTCTGTTGATCAAGAGCGCGCTCTTTTGA 377
2b 181 TCTCGAAGCTGAATGATGCGCCCTTTCTGTTGATCAAGAGCGCGCTCTTTTGA 240
2y 378 GGAATGATGATGTTCACTGGAAGAAATGGGACATTAGTTCAAGTAGCACTATATC 437
2b 241 GGAATGATGATGTTCACTGGAAGAAATGGGACATTAGTTCAAGTAGCACTATATC 300
2y 438 AGAAACATGTTCAACCAATGCGAAAGTGGGTAAACAGACATATGAACAGAAATTTA 457
2b 301 AGAAACATGTTCAACCAATGCGAAAGTGGGTAAACAGACATATGAACAGAAATTTA 360
2y 498 TTATGACATGATGATGAAAGGACGCGCAAGAAAGGCGGACAGACATGTTGATTC 557
2b 361 TTATGACATGATGATGAAAGGACGCGCAAGAAAGGCGGACAGACATGTTGATTC 420
2y 558 CTAGACCTGTTCCAAATTTGTTTAAAGACCTTTTAAAGTTGGCTGAATTTGAGCAGA 617
2b 421 CTAGACCTGTTCCAAATTTGTTTAAAGACCTTTTAAAGTTGGCTGAATTTGAGCAGA 480
2y 618 GTTCAAGACATAGAAACCAACTTATCAAGAAATTTCTTTACGTGAGACCTATTTA 677
2b 481 GTTCAAGACATAGAAACCAACTTATCAAGAAATTTCTTTACGTGAGACCTATTTA 540
2y 678 TCTGGAAGTGAACATCTGTTTGGGCGCAAGAAACAGACCTGTTGTTAGCAT 737
2b 541 TCTGGAAGTGAACATCTGTTTGGGCGCAAGAAACAGACCTGTTGTTAGCAT 600
2y 738 AAAAAGATTTATTTACCCCTTCAACCAACATTTGCCAATTAAGAAATTTCTGTTAGCT 797
2b 601 AAAAAGATTTATTTACCCCTTCAACCAACATTTGCCAATTAAGAAATTTCTGTTAGCT 660
2y 798 CTTCGAATTTTATGATGATGATGTCACAAACAGTCTTATTTGTTATATTTTGA 857
2b 661 CTTCGAATTTTATGATGATGATGTCACAAACAGTCTTATTTGTTATATTTTGA 716
2y 858 ATATGATTTTACCTTGAATATCCCTTTTATTAATAATATGATGAGAAATCCCTTT 917
2b 717 ATATGATTTTACCTTGAATATCCCTTTTATTAATAATATGATGAGAAATCCCTTT 776
2y 918 AC 919
2b 777 TC 778

RESULT 6
AK043247 2356 bp mRNA linear HTC 19-SEP-2003
LOCUS AK043247 Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length
DEFINITION CEROID-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLNS PROTEIN) [Homo
sapiens], full insert sequence.
ACCESSION AK043247
VERSION AK043247.1 GI:26089577
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
TITLE High-efficiency full-length cDNA cloning
JOURNAL Weich. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
TITLE High-efficiency full-length cDNA cloning
JOURNAL Weich. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 204939374

```

```

PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
JOURNAL sequencing pipeline with 384 multichipillary sequencer
MEDLINE 20530913
PUBMED 11076861
4
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akiyama, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagata, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohata, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takatsu-Ahahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshinori Hayashizaki. The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
LOCATION/Qualifiers
FEATURES
source
1. .2356
/organism="Mus musculus"
/mol_type="mRNA"
/script="5'UTR/5'"
/db_xref="FANTOM_DB:A730075N08"
/db_xref="MGI:2408405"
/db_xref="taxon:10090"
/clone="A730075N08"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="7 days neonate"
1. .2356
/note="similar to CEROID-LIPOFUSCINOSIS NEURONAL PROTEIN 5
(CNS PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence:
FASTA, 73.9%ID, 86.7%length, match=1041)"
ORIGIN
Query Match 70.9%; Score 671.8; DB 11; Length 2356;

```

Best Local Similarity 81.8%; Pred. No. 3.9e-142;  
Matches 775; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

```

QY 1 CGGGTCTCGGAGCATCCCTCCCGGCGCACTGCGCGGCTGCGCTTGAAGGCTTACTTC 63
DB 130 CTGGGCGCGTCCCGACGCTCTGGGCAACGCTGGCCGGTCCCTCAAGCGCTTCTTTC 139
QY 61 GGTCCAAAACCTGATCTTATTTGCAAGTAAATTAATTTCTGTCACATGGCTCACCT 120
DB 130 GCTCCGAAAGACAGATCCCTACTGCTCAAGCTAATTAATTTCTGTCACATGGCTCACCT 249
QY 121 ATCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 130
DB 250 ATCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 309
QY 181 GAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
DB 310 GAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 369
QY 241 AGAAGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
DB 370 AGAAGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 429
QY 301 AACTGTACATTTCCCGACATCTCGACATGAATGATGCGCTTCTGCTGATACAAAGCC 360
DB 430 AACTGTACATTTCCCGACATCTCGACATGAATGATGCGCTTCTGCTGATACAAAGCC 489
QY 361 GCTGCTCTCTTTTGAAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420
DB 480 GCAAGCTCTCTTTTGAAGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 549
QY 421 CAAGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
DB 550 GTCGTGCAACATTCGGAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 609
QY 481 AATGAAACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540
DB 610 AATGAAACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 669
QY 541 GAGACATGTTGATTTCTACGATCTGCAATTAATTAATTAATTAATTAATTAATTAAT 600
DB 670 CAGAGTGTGTTGAGTCTTACGATCTGCAATTAATTAATTAATTAATTAATTAATTAAT 729
QY 601 GCTGAATTTGAGCAGAGTTCAGAAACATTAATTAATTAATTAATTAATTAATTTCTTAC 660
DB 730 GCTGAATTTGAGCAGAGTTCAGAAACATTAATTAATTAATTAATTAATTAATTTCTTAC 789
QY 661 AGTGAAGAACCTACTTATCTGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 720
DB 790 AGTGAAGAACCTACTTATCTGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 849
QY 721 ACTCTTGTTAGCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 780
DB 850 ACTCTTGTTAGCCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 909
QY 781 GAATTTCTGTTGAGTCTCTTCAATTTTGAATGAGTATGTCACAAACAGTTCTAT 840
DB 910 GAATTTCTGTTGAGTCTCTTCAATTTTGAATGAGTATGTCACAAACAGTTCTAT 969
QY 841 TTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 900
DB 970 TTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1029
QY 901 TATGAAGAAATCCCTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 947
DB 1030 TATGAAGAAATCCCTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1076

```

## RESULT 7

BU621797/c

LOCUS BU621797 689 bp mRNA linear EST 23-SEP-2002  
DEFINITION UI-H-FL1-bgd-i-07-0-UI-s1 NCI CGAP\_Fl1 Homo sapiens cDNA clone  
UI-H-FL1-bgd-i-07-0-UI-3', mRNA sequence.

ACCESSION BU621797  
VERSION BU621797.1 GI:23288012  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapds-remail.nih.gov  
Tissue Procurement: James Martin

CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
Seq primer: M13 FORWARD  
POLYA=yes.

## FEATURES

Source  
1..689  
/location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-FL1-bgd-i-07-0-UI"  
/issue\_type="Cell lines"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Chondrosarcoma; Vector: pRT3-Pac (pharmacia) with a modified polylinker, Site 1: EcoR I; Site 2: Not I; NCI CGAP\_Fl1 is a normalized cDNA library derived from a pool of mRNA obtained from 4 cell lines from grade III chondrosarcoma tissues. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GAGTGGCTG. The cell lines were provided by Dr. James Martin from the University of Iowa.  
TAG\_TISSUE=Human Chondrosarcoma Grade 3 cell line mix  
TAG\_LIB=UI-H-FL1  
TAG\_SEQ=GAGTGGCTG"

## ORIGIN

Query Match 70.7%; Score 670.6; DB 13; Length 689;  
Best Local Similarity 99.4%; Pred. No. 8.3e-142;  
Matches 673; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 234 TGGATTGAGAAAGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 293
DB 689 TGGATTGAGAAAGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 630
QY 294 ACTTGGCACTGTACATTTCCCATCTCCGACCTGAATGATGCGCTTCTGCTGTTAA 353
DB 629 ACTTGGCACTGTACATTTCCCATCTCCGACCTGAATGATGCGCTTCTGCTGTTAA 570
QY 354 TCAAGGCGCTGCTGCTTTTGAAGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 413
DB 569 TCAAGGCGCTGCTGCTTTTGAAGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 510
QY 414 ATTAGTCAAGTACACTATATCAGAAACATGTTCAACCAATGCAAAAGTGGTGAA 473
DB 509 ATTAGTCAAGTACACTATATCAGAAACATGTTCAACCAATGCAAAAGTGGTGAA 450

```

Y 474 ACAGACATGAAACAGAAATTATATGACATGAAATGTAAGACGACGACGAGAAA 533  
 b 449 ACAGACATGAAACAGAAATTATATGACATGAAATGTAAGACGACGACGAGAAA 393  
 Y 534 GGGGGCAGACATGGTTGATTCCTACAGACTGTTCCAAATTTGTTAGGACCTTTAA 593  
 b 389 GGGGGCAGACATGGTTGATTCCTACAGACTGTTCCAAATTTGTTAGGACCTTTAA 333  
 Y 594 CAAAGTGGCTGAATTTGGAGCAGAGTTCAAGAACATGAAACCACTATACAGAAATTT 653  
 b 329 CAAAGTGGCTGAATTTGGAGCAGAGTTCAAGAACATGAAACCACTATACAGAAATTT 273  
 Y 654 TCTTTACAGGAGAACCTACTATCTCGGGAATGAAACATCTGTTTGGGCGAACAG 713  
 b 269 TCTTTACAGGAGAACCTACTATCTCGGGAATGAAACATCTGTTTGGGCGAACAG 213  
 Y 714 AACACAGACTCTGTTGATTCCTACAGACTGTTCCAAATTTGTTAGGACCTTTAA 773  
 b 209 AACACAGACTCTGTTGATTCCTACAGACTGTTCCAAATTTGTTAGGACCTTTAA 153  
 Y 774 AACACAGACTCTGTTGATTCCTACAGACTGTTCCAAATTTGTTAGGACCTTTAA 833  
 b 149 AACACAGACTCTGTTGATTCCTACAGACTGTTCCAAATTTGTTAGGACCTTTAA 90  
 Y 834 GTTCTATTTGTTTAAATTTGAAATATGTTTAACTATGAAATTCCTTTATTTAA 893  
 b 89 GTTCTATTTGTTTAAATTTGAAATATGTTTAACTATGAAATTCCTTTATTTAA 30  
 Y 894 AATTAACATATGAGAAA 910  
 b 29 AATTAACATATGAGAAA 13  
 RESULT 8  
 AK085741 2433 bp mRNA linear HTC 20-SEP-2003  
 ACUS AK085741  
 DEFINITION Mus musculus 10 days lactation, adult female mammary gland cDNA,  
 RIKEN full-length enriched library, clone:DJ30033P03  
 Product: similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5  
 PROTEIN) [Homo sapiens], full insert sequence.  
 ACCESSION AK085741  
 VERSION AK085741.1 GI:26102938  
 EYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1  
 Carninci, P. and Hayashizaki, Y.  
 TITLE High efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636  
 REFERENCE 2  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159  
 REFERENCE 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Konno, H., Akiyama, J., Ishii, K., Katsunari, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, H., Harada, A.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuoka, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913

PUBMED 11076861  
 REFERENCE 4  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE 5  
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 REFERENCE 6  
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kawaga, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Konda, M.,  
 Koyas, S., Kuribara, C., Matsuyama, T., Miyazaki, A., Morita, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ono, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-Apr-2002) Yoshihide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan [E-mail: genome-res@gs.c.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
 Fax: 81-45-503-9216]  
 COMMENT cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 URL: http://fantom.gsc.riken.go.jp/.  
 FEATURES  
 source location/Qualifiers  
 1..2433  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:DJ30033P03"  
 /db\_xref="MGI:2422937"  
 /db\_xref="taxon:10090"  
 /clone="DJ30033P03"  
 /sex="female"  
 /tissue\_type="mammary gland"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="10 days lactation, adult"  
 1..2433  
 /note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5  
 (CLN5 PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence:  
 FASTA, 73.9%ID, 86.7%length, match=1041)"  
 ORIGIN  
 Query Match 70.7% Score 670.2 DB 11: Length 2433;  
 Best Local Similarity 81.7% Pred. No. 8.9e-142;  
 Matches 774; Conservative 0; Mismatches 173; Indels 0; Gaps 0;  
 Y 1 CGGGTCTGGGGGATCCCTCCGCGGCGACACTGGCGGTCCTACAGGCGCTTGACTTC 60  
 Db 207 CTGGGCGGCTCCCGACGCTTGGGCGAGCGTGGCGGTCCTACAGGCGCTTCTTTTC 266  
 Y 61 CGTCCAAACCTGATCTTATTTGCAAGCTAAGTATCTTTGTGTCGAACCTGGCTACCT 120  
 Db 267 CGTCCGAAGACAGATCCCTCACTGCTCAAGCTAAGTATCTTTGTGTCCTACCGGCTGCGCC 326  
 Y 121 ATCCAGTTATGAGAGGAGATGATGATCACTGAAGTTTTCGATTCACAGCCCAAGTATG 180





CDS

```

/sex="male"
/issue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
<1..971
/notes="unnamed protein product; putative
similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CNLS
PROTEIN) (Homo sapiens) (SWISSPROT|075503, evidence:
FASTA, 73.9%ID, 86.7%length, match=1041)"
/codon_start=3
/protein_id="BAC27797.1"
/db_xref="GI:26328113"
/translation="LGLATILGASPTSGORPVYKHSFRPKTPDYCAKYECPPG
SPIPMDNDVIEVLRIQAPIMERKYGDIQHPEDMDAYGRSTLGRKYTERKYL
FOLGNCITPHIRPDKAPFPMQNSAACFFRIGIDDKRKENGTTLSVATISGTFNKYA
EWMKQDNBTGIIYETWTWRAGPQQAQWESIDCSNPLRTYKLAEPFETEKIET
NYTKIFLYSGEPPIYLGNETSI FGRKNGKTLALIKFYGPRPLSTKDFIMFLKIF
DTVIHQFLYLFYFEYFWFLPMKPPFKITIEETPLTRHTFTDL"

```

ORIGIN

Query Match 70.5%; Score 668.6; DB 11; Length 2247;

Best Local Similarity 81.6%; Pred. No. 2.1e-141;

Matches 773; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

```

1 CGGGCTCGGGCATCCCTCCCGGCGCACTGCGCGGCTTACAGCGCTTGACTTC 60
21 CTGGGCGCGTCCCGGCGAGCTCTGGGCAACGCTGCGGCTCATCAAGCACTTCTTTC 80
61 CGTCCAAACCTGATCTTATTTGCTAAGTATATCTTCTGCTCACTGGCTCACT 120
81 CGTCCGAAGACAGATCCCTACTGTCAAGCTATATCTTCTGCTCACTGGCTGCCC 140
121 ATCCCACTTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATG 160
141 ATCCCACTTATGAGGAGCAATGATGATGATGATGATGATGATGATGATGATG 200
181 GAATTTAATATGAGAGCTCTGCGGCACTGCGGCTTATGATGATGATGATGATG 240
201 GAATTTAATATGAGAGCTCTGCGGCACTTAACTTATGATGATGATGATGATG 260
241 AGAATGATCATTAACTGCGCAAGAACTATCAATGATGATGATGATGATGATG 300
261 AGAGGACACATGACAGCAAGAACTATCAATGATGATGATGATGATGATGATG 320
301 AACTGTACATTTTCCCATCTCCGACTGGAATGATGATGATGATGATGATGATG 360
321 AACTGTACATTTTCCCATCTCCGACTGGAATGATGATGATGATGATGATGATG 380
361 GCTGCGCTGCTTTTGGGAGGATGATGATGATGATGATGATGATGATGATGATG 420
381 GCACTGCTGCTTTTGGGAGGATGATGATGATGATGATGATGATGATGATGATG 440
421 CAATGATCAATATATGAGGAACTATGATGATGATGATGATGATGATGATGATG 460
441 GTGCTTCAACCAATATCGGAAACATTTAAAGAACTGCGGAGTGGTGAAGGAG 500
481 AATGAAACAGGAATTTTATGAGACATGGAATGGAATGGAATGGAATGGAATG 540
501 AATGAAACAGGAATTTTATGAGACATGGAATGGAATGGAATGGAATGGAATG 560
541 GAGCAGATGCTTTGATGCTGAGACTGTTCAAAATTTGTTAAGGACTTTAAGATG 600
561 CAGCTGCTGCTTTGATGCTGAGACTGTTCAAAATTTGTTAAGGACTTTAAGATG 620
601 GCTGAATTTGAGACAGATTTCAAGAACATGGAACCAATATATCAAGATATTTCTTAC 660
621 GCTGAATTTGAGACAGATTTCAAGAACATGGAACCAATATATCAAGATATTTCTTAC 680
661 AGTGAAGAACTTACTTATCTGGGAAATGAAACATCTTTTGGGCAACAGGAAACAG 720
681 AGTGAAGAACTTACTTATCTGGGAAATGAAACATCTTTTGGGCAACAGGAAACAG 740
721 ACTCTGTTTAAAGCATTAAGAAAGATTTTATTAACCCCTTCAACCACTTTGCCAATAAA 780

```

```

Db 741 ACTCTGCTTTGAGCATTAAGAAAGATTTTATTAACCCCTTCAACCACTTTGCCAATAAA 800
Qy 781 GAATTTCTGTTGAGCTCTTTCGAAATTTTATGATGATGATGATGATGATGATGATGATG 840
Db 801 GATTTCTGATGATGATTTCTTGAAGAAATTTTATGATGATGATGATGATGATGATGATGATG 860
Qy 841 TTGTTTATTAATTTTGAATTTTGGTTTATGATGATGATGATGATGATGATGATGATGATG 900
Db 861 TTGTTTATTAATTTTGAATTTTGGTTTATGATGATGATGATGATGATGATGATGATGATG 920
Qy 901 TATGAAGAAATCCCTTATCTATCAGAAACAAACTCTCTGATTT 947
Db 921 TAGGAAGAAACCCGTTACTTACTTACCAGATACAAATTTACCACTT 967

```

RESULT 10

BX370641

LOCUS

DEFINITION

BX370641 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED

ACCESSION

BX370641

VERSION

BX370641.1 GI:30459737

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens (human)

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 908)

AUTHORS

I.J.W.B., Gruber,C., Jessee,J. and Polayes,D.

TITLE

Full-length cDNA libraries and normalization

JOURNAL

Unpublished (2001)

COMMENT

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 7238.f For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0BA053ZH03 CS05016.1cluster=7238.f.

Contact: Peng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paraday Avenue Genoscope sequence ID: CS0BA053ZH03\_CS05016.1.

Location/Qualifiers

1..908

/organism="Homo sapiens"

/mol\_type="RNA"

/db\_xref="taxon:9606"

/clone="CS0D004YP03"

/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"

/cell\_line="RAMOS CELL LINE"

/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT

25-NORMALIZED"

/note="Left strand cDNA was primed with a NotI-oligo (dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 68.1%; Score 645.4; DB 13; Length 908;

Best Local Similarity 98.2%; Pred. No. 4.2e-136;

Matches 663; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

```

Qy 274 GAATGATGAACTTTTTCGAACCTGCAACTGATGATGATGATGATGATGATGATGATG 333
Db 1 GATGATGATGAACTTTTTCGAACCTGCAACTGATGATGATGATGATGATGATGATGATG 59
Qy 334 GATGATGATGAACTTTTTCGAACCTGCAACTGATGATGATGATGATGATGATGATGATG 393
Db 60 GATGATGATGAACTTTTTCGAACCTGCAACTGATGATGATGATGATGATGATGATGATG 119
Qy 394 CACTGAGAGAGAAATGAGCACTAGTTCAAGTATGCAACTATATCAGAGAAACATGTTCAAC 453

```

Db 120 CACTGGAAGGAAATGGGACATTAGTCAAGTGAACATATATCGAGAAACATGTTTAC 179  
Qy 454 CAATGCGCAAGTGGGTGTAACAGAGCAATGAACAGGATTTATATGACATGAT 513  
Db 180 CAAATGCGCAAGTGGGTGTAACAGAGCAATGAACAGGATTTATATGACATGAT 239  
Qy 514 GTAAAGCCAGCCAGAAAGGGGCGAGAGACATGTTGATTTCTTACAGCTGTCCAA 573  
Db 240 GTAAAGCCAGCCAGAAAGGGGCGAGAGACATGTTGATTTCTTACAGCTGTCCAA 239  
Qy 574 TTTGTTTAAAGACCTTTTAAAGATGTTGCTGAATTTGAGCAGATTTAAAGACATGAA 633  
Db 300 TTTGTTTAAAGACCTTTTAAAGATGTTGCTGAATTTGAGCAGATTTAAAGACATGAA 339  
Qy 634 ACCAATCTATACAGATATTTCTTTACAGTGAAGAACCTACTTATCTGGGAAATGAACA 693  
Db 360 ACCAATCTATACAGATATTTCTTTACAGTGAAGAACCTACTTATCTGGGAAATGAACA 419  
Qy 694 TCTGTTTGGGCGACAGAAACAAAGACTCTTGGTTAGCCATTAAGATTTTATATAC 753  
Db 420 TCTGTTTGGGCGACAGAAACAAAGACTCTTGGTTAGCCATTAAGATTTTATATAC 479  
Qy 754 CCCTTCAACACCATTTGCGAATTAAGATTTCTGTTGAGTCTCTTCAATTTTATAT 813  
Db 480 CCCTTCAACACCATTTGCGAATTAAGATTTCTGTTGAGTCTCTTCAATTTTATAT 539  
Qy 814 GCAGTATGTCACAAACAGTTCTATTTGTTTATTTATTTGAATTTGTTTATACCT 873  
Db 540 GCAGTATGTCACAAACAGTTCTATTTGTTTATTTATTTGAATTTGTTTATACCT 599  
Qy 874 ATGAATTTCCCTTTATTTAAATTAATATGAAGAAATCCCTTATCTATCAGAAACAA 933  
Db 600 ATGAATTTCCCTTTATTTAAATTAATATGAAGAAATCCCTTATCTATCAGAAACAA 659  
Qy 934 AACCTCTCTGTTTA 948  
Db 660 AACCTCTCTGTTTA 674

RESULT 11  
CA488543 924 bp mRNA linear EST 14-NOV-2002  
LOCUS CA488543  
DEFINITION AGENCOURT\_10808864 MAPCL Homo sapiens cDNA clone IMAGE:6720309 5',  
mRNA sequence.  
ACCESSION CA488543  
VERSION CA488543.1 GI:24950702  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 924)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apds-remail.nih.gov  
Tissue Procurement: Kristi A. Egland, Ira Pastan  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LLM14279 row: j column: 21  
High quality sequence stop: 637.  
Location/Qualifiers  
1..924  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6720309"

/cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,  
hTERT-HME1, INCaP"  
/lab\_host="EMD108"  
/clone\_lib="MAPCL"  
/note="Vector: pCMV-SPORT6; Site 1: EcoRV; Site 2: Not I;  
Subcloned with brain, liver, lung, kidney and muscle.  
Directionally cloned. Priming method: oligo-dT. Average  
insert size: 1800 bp. Library amplification: 26,000 fold.  
Kristi A. Egland, James J. Vincent, Robert Strausberg,  
Bungkok Lee & Ira Pastan. Discovery of new breast  
cancer genes encoding membrane and secreted proteins.  
Manuscript submitted."

Query Match 67.8%; Score 643.2; DB 14; Length 924;  
Best Local Similarity 96.6%; Pred. No. 1.3e-135;  
Matches 657; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 1 CGGGTCTCGGCAATCCCTCCGCGCACTGCGCGGCTCCCAAGCGCTTGACTTC 60  
Db 106 CGGGTCTCGGCAATCCCTCCGCGCACTGCGCGGCTCCCAAGCGCTTGACTTC 165  
Qy 61 CGTCCAAACCTGATCTTATTTGCAAGCTATATCTTTCTGCACTGCTCACT 120  
Db 166 CGTCCAAACCTGATCTTATTTGCAAGCTATATCTTTCTGCACTGCTCACT 225  
Qy 121 ATCCAGTTATGAGGAGTATGATGATGATGATGATGATGATGATGATGATG 180  
Db 226 ATCCAGTTATGAGGAGTATGATGATGATGATGATGATGATGATGATG 285  
Qy 181 GAATTTAAATATGAGACCTCTCTGGAACCTTGAATTAATGATGATGATG 240  
Db 286 GAATTTAAATATGAGACCTCTCTGGAACCTTGAATTAATGATGATGATG 345  
Qy 241 AGAAGTACATTAATCTGCAAGAACTACACATGATGATGATGATGATGATG 300  
Db 346 AGAAGTACATTAATCTGCAAGAACTACACATGATGATGATGATGATG 405  
Qy 301 AACTGTACATTTCCCAATCTCCGACCTGAATGATGATGATGATGATGATG 360  
Db 406 AACTGTACATTTCCCAATCTCCGACCTGAATGATGATGATGATGATG 465  
Qy 361 GCTGCTGCTTTTGTGAGGAAATGATGATGATGATGATGATGATGATGAT 420  
Db 466 GCTGCTGCTTTTGTGAGGAAATGATGATGATGATGATGATGATGATGAT 525  
Qy 421 CAAGTACACATTAATCTGCAAGAACTACACATGATGATGATGATGATG 480  
Db 526 CAAGTACACATTAATCTGCAAGAACTACACATGATGATGATGATGATG 585  
Qy 481 AATGAAACAGAAATTTATTTGAGACATGATGATGATGATGATGATGATG 540  
Db 586 AATGAAACAGAAATTTATTTGAGACATGATGATGATGATGATGATGATG 645  
Qy 541 GAGACATGTTTATTTCTGCACTGTTCCAAATTTGTTAAGACCTTTAA 600  
Db 646 GAGACATGTTTATTTCTGCACTGTTCCAAATTTGTTAAGACCTTTAA 705  
Qy 601 GCGGAATTTGAGCAGAGTTCAGAACATGAAACCAATGATCAAGAAATTT 660  
Db 706 GCGGAATTTGAGCAGAGTTCAGAACATGAAACCAATGATCAAGAAATTT 765  
Qy 661 AGTGAAGAACTTACTTACT 680  
Db 766 ACAGTGGAGAAACCTTACT 785

RESULT 12  
BI090566 869 bp mRNA linear EST 20-JUN-2001  
LOCUS BI090566  
DEFINITION 602855673F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:4996891 5',  
mRNA sequence.  
ACCESSION BI090566

ERSION BI090566.1 GI:14508896  
 EYMORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 869)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strauberg, Ph.D.  
 Email: c9apbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM1023 row: a column: 20  
 High quality sequence stop: 843.  
 Location/Qualifiers  
 1..869  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:498891"  
 /cell\_line="MGC36"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_10"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.5 Kb. Library prepared by Life  
 Technologies."

RIGIN  
 Query Match 66.4%; Score 629.6; DB 12; Length 869;  
 Best Local Similarity 96.8%; Pred. No. 1,6e-132;  
 Matches 706; Conservative 0; Mismatches 15; Indels 8; Gaps 5;

```

Y 1 CGGGGTCTCGGGGATCCCTCCCGGCGCACTGGCCGCTCCCTACAGCGCTTGGACTTC 60
b 142 CGGGGTCTCGGGGATCCCTCCCGGCGCACTGGCCGCTCCCTACAGCGCTTGGACTTC 201
Y 61 CGTCCAAAACCTGATCTTATGTCAGCTAAGTATATCTTCTGTCCAACTGGCTCACT 123
b 202 CGTCCAAAACCTGATCTTATGTCAGCTAAGTATATCTTCTGTCCAACTGGCTCACT 261
Y 121 ATCCGAGTATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 183
b 262 ATCCGAGTATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 321
Y 181 GAATTTAAATATGAGAGACTCTCTGGGACACTTGAATAATATGATGATGATGATGATG 240
b 322 GAATTTAAATATGAGAGACTCTCTGGGACACTTGAATAATATGATGATGATGATGATG 381
Y 241 AGAAGTACATTAACTGGCAAGAACTACAAATGGAATGGTAACTTTTCCACTTGGC 300
b 382 AGAAGTACATTAACTGGCAAGAACTACAAATGGAATGGTAACTTTTCCACTTGGC 441
Y 301 AACTGTACATTTCCCATCTCCGACTGGAATGGAATGGAATGGAATGGAATGGAATG 360
b 442 AACTGTACATTTCCCATCTCCGACTGGAATGGAATGGAATGGAATGGAATGGAATG 501
Y 361 GCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
b 502 GCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 561
Y 421 CAAGTACGATCTATTCAGGAAACATGTTCAACCAATGGCAAGTGGTGAAGCAAGAC 480
b 562 CAAGTACGATCTATTCAGGAAACATGTTCAACCAATGGC-AAAGTGGTGAAGCAAGAC 620
Y 481 AATGAACAGCAATTTATATGAGACATGATGATGATGATGATGATGATGATGATGATG 538
  
```

Db 621 AATGAACAGCAATTTATATGAGACATGATGATGATGATGATGATGATGATGATGATG 680

QY 539 CAGAGACATGTTGTTGATTTCTTACAGCTGTCCAAATTTTGTGTAGAGACCTTTAAGAGT 598

Db 681 AGAGACATGTTGTTGATTTCTTACAGCTGTCCAAATTTTGTGTAGAGACCTTTAAGAGT 740

QY 599 TGGCTGAA-TTGGAGCAGAGTTCAAGAA-TCGAACCACTATACAGAAATATTTCT 656

Db 741 TGGCTGAACTGGAGCAGAGTTCAAGAACTATGAAACCAATATACAGAAATATTTCT 800

QY 657 TTAAGTGG-AGAACTACTTATCTGGGAAA-TGAACATCTGTTTGGGCGCAACAG 713

Db 801 TTAAGTGGAGAAACCTACTTATCTGGGAAAATGAACACATCTTTAGGGCGCAACAG 860

QY 714 AAACAGAC 722

Db 861 AAACAGAC 869

RESULT 13  
 B1223533  
 LOCUS 787 bp mRNA linear EST 11-JUN-2001  
 DEFINITION 602941923F1 NIH\_MGC\_12 Homo sapiens CDNA clone IMAGE:5105093 5',  
 mRNA sequence.  
 B1223533  
 VERSION B1223533.1 GI:14676977  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 787)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strauberg, Ph.D.  
 Email: c9apbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM1253 row: n column: 06  
 High quality sequence stop: 762.  
 Location/Qualifiers  
 1..787  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5105093"  
 /tissue\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_12"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1; NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.4 Kb. Library prepared by Life  
 Technologies."

ORIGIN  
 Query Match 63.4%; Score 601; DB 12; Length 787;  
 Best Local Similarity 98.5%; Pred. No. 5.1e-126;  
 Matches 659; Conservative 0; Mismatches 5; Indels 5; Gaps 5;

```

QY 1 CGGGCTCGGGGATCCCTCCCGGCGCACTGGCCGCTCCCTACAGCGCTTGGACTTC 60
Db 116 CGGGCTCGGGGATCCCTCCCGGCGCACTGGCCGCGCTTACAGCGCTTGGACTTC 175
QY 61 CGTCCAAAACCTGATCTTATGTCAGCTAAGTATATCTTCTGTCCAACTGGCTCACT 120
Db 176 CGTCCAAAACCTGATCTTATGTCAGCTAAGTATATCTTCTGTCCAACTGGCTCACT 235
  
```

```
OY 121 ATCCGATTGAGAGGCTGATGATGACATGTAAGTTTTCATTAACAAGCCCAAGTATG 130
DB 236 ATCCGATTGAGAGGCTGATGATGACATGTAAGTTTTCATTAACAAGCCCAAGTATG 235
OY 181 GAATTTAAATNAGAGACCTCTGGGACACTTGAAATTAAGCATGATGCAATGATTC 240
DB 236 GAATTTAAATNAGAGACCTCTGGGACACTTGAAATTAAGCATGATGCAATGATTC 355
OY 241 AGAATGATTAATGAGAGACACTTGCAACAATGAGATGATGATGATGATGATGATGATG 300
DB 356 AGAATGATTAATGAGAGACACTTGCAACAATGAGATGATGATGATGATGATGATGATG 415
OY 301 AACTGATGATTTCCCATCTCCGACCTGAAATGAGATGATGATGATGATGATGATGATG 360
DB 416 AACTGATGATTTCCCATCTCCGACCTGAAATGAGATGATGATGATGATGATGATGATG 475
OY 361 GCTGCTCTCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 476 GCTGCTCTCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATG 535
OY 421 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGCAAGTGGGTGAAACAGGAC 480
DB 536 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGCAAGTGGGTGAAACAGGAC 595
OY 481 AATGAAACAGAAATTTATATGAGACATGAAATGTAAGCAAGCCCAAGAAAGGAGCA 540
DB 596 AATGAAACAGAAATTTATATGAGACATGAAATGTAAGCAAGCCCAAGAAAGGAGCA 654
OY 541 GAGACATGTTGATCTCTAGACCTGTTCCAAATTTGTTT-AAGACCTTTTAACTT 599
DB 655 GAGACATGTTGATCTCTAGACCTGTTCCAAATTTGTTT-AAGACCTTTTAACTT 714
OY 600 GGCTGA-ATTGAGAGCAAGTTCAAGACATA-GAAACCACTATACAG-AATATTTCT 656
DB 715 GGCTGA-ATTGAGAGCAAGTTCAAGACATA-GAAACCACTATACAG-AATATTTCT 774
OY 657 TTACAGTGG 665
DB 775 CTACAGTGG 783
```

```
RESULT 14
LOCUS CB215456 641 bp mRNA linear EST 06-FEB-2003
DEFINITION NISC mp06c07.y1 NICHD_HS_Ut1 Homo sapiens CDNA IMAGE:5937013
VERSION CB215456
KEYWORDS CB215456.1 GI:28263648
SOURCE EST.
ORGANISM Homo sapiens (human)
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-rc@mail.nih.gov
CDNA Library Preparation:
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LINL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
info@image.jnl.gov
Plate: LHM1163 row: E column: 14
Seq primer: M13R1 reverse primer (ABI).
Location/Qualifiers
1..641
/organism="Homo sapiens"
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5937013"
/sex="female"
/tissue_type="normal endometrium, late proliferative
phase, cycle day 13"
/lab_host="DH10B (TI-resistant)"
/clone_id="NICHD HS Ut1"
/note="Organ: uterus; Vector: pCMV-Sport6.1.cdb (Regen,
Invitrogen Corporation); Site 1: Not; Site 2: EcorV;
Cloned unidirectionally from microquantity amounts of mRNA
from normal endometrial tissue (late proliferative phase,
cycle day 13). Average insert size 1.9 kb. Library
constructed by Regen (Invitrogen Corporation)."
```

```
Query Match 63.3%; Score 600; DB 14; Length 641;
Best Local Similarity 100.0%; Pred. No. 8-8e-126;
Matches 600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 CGGGTCTCGGCAATCCCTCCCGGCGGCGGCTGCAAGCGCTTGACTTC 60
DB 42 CGGGTCTCGGCAATCCCTCCCGGCGGCGGCTGCAAGCGCTTGACTTC 101
OY 61 CGTCAAAACCTGATCTTATTTGCAAGCTATATCTTCTGCTCACTGCTCACT 120
DB 102 CGTCAAAACCTGATCTTATTTGCAAGCTATATCTTCTGCTCACTGCTCACT 161
OY 121 ATCCAGTTATGAGAGGATGATGATGATGATGATGATGATGATGATGATG 180
DB 162 ATCCAGTTATGAGAGGATGATGATGATGATGATGATGATGATGATGATG 221
OY 181 GAATTTAAATATGAGAGCTCTCTGGGACACTTGAATTAATGATGATGATGATG 240
DB 222 GAATTTAAATATGAGAGCTCTCTGGGACACTTGAATTAATGATGATGATGATG 281
OY 241 AGAATGATTAATGAGAGCTCTCTGGGACACTTGAATTAATGATGATGATGATG 300
DB 282 AGAATGATTAATGAGAGCTCTCTGGGACACTTGAATTAATGATGATGATGATG 341
OY 301 AACTGTACATTTCCCATCTCCGACCTGAATGATGATGATGATGATGATGATG 360
DB 342 AACTGTACATTTCCCATCTCCGACCTGAATGATGATGATGATGATGATGATG 401
OY 361 GCTGCCCTGCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATG 420
DB 402 GCTGCCCTGCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATG 461
OY 421 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGCAAGTGGGTGAAACAGGAC 480
DB 462 CAAGTAGCACTATATCAAGAAACATGTTCAACCAATGCAAGTGGGTGAAACAGGAC 521
OY 481 AATGAAACAGAAATTTATATGAGACATGAAATGTAAGCAAGCCCAAGAAAGGAGCA 540
DB 522 AATGAAACAGAAATTTATATGAGACATGAAATGTAAGCAAGCCCAAGAAAGGAGCA 581
OY 541 GAGACATGTTGATCTCTAGACCTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 600
DB 582 GAGACATGTTGATCTCTAGACCTGTTCCAAATTTGTTAAGACCTTTAACAAGTTG 641
```

```
RESULT 15
LOCUS BF982158 1098 bp mRNA linear EST 23-JAN-2001
DEFINITION 602308976F1 NIH_MGC_88 Homo sapiens CDNA clone IMAGE:4400298 5',
VERSION BF982158
KEYWORDS BF982158.1 GI:12384970
SOURCE EST.
ORGANISM Homo sapiens (human)
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE 1 (bases 1 to 1098)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LHAM10105 row: 9 column: 19  
High quality sequence stop: 693.  
Location/Qualifiers  
1..1098

FEATURES  
source  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:440298"  
/tissue\_type="duodenal adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: small intestine; Vector: pCMV-SPORT6;  
Site\_1: Not1; Site\_2: Sal1; Cloned unidirectionally;  
oligo-dr primed. Average insert size 1.767 kb. Library  
enriched for full-length clones and constructed by Life  
Technologies. Note: this is a NIH\_MGC Library."

Query Match 59.3%; Score 562.4; DB 10; Length 1098;  
Best Local Similarity 88.2%; Pred. No. 2.9e-117;  
Matches 660; Conservative 0; Mismatches 76; Indels 12; Gaps 4;

1 CGGGCTCGGGGCAATCCCTCCCGGGGCGCACTGCGCGCTTACAAAGCGCTTGACTTC 60  
2 CGGGCTCGGGGCAATCCCTCCCGGGGCGCACTGCGCGCTTACAAAGCGCTTGACTTC 184  
61 CGTCCAAACCTGATCCTTATGTCAGCTAAGTATACCTTCTGTCACACTGGCTCACT 120  
185 CGTCCAAACCTGATCCTTATGTCAGCTAAGTATACCTTCTGTCACACTGGCTCACT 244  
121 ATCCGAGTTATGAGGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
245 ATCCGAGTTATGAGGGTATGATGATGATGATGATGATGATGATGATGATGATGAT 304  
181 GAATTTAAATATGAGAGCTCCCTGGGCACTTGAAATTTATGATGATGATGATGAT 240  
305 GAATTTAAATATGAGAGCTCCCTGGGCACTTGAAATTTATGATGATGATGATGAT 364  
241 AGAAGTACATTAATGAGCAAGAACTACACATGAAATGATGAACTTTCCAACTTGGC 300  
365 AGAAGTACATTAATGAGCAAGAACTACACATGAAATGATGAACTTTCCAACTTGGC 424  
301 AACTGTCATTTCCCACTCCGACCTGAAATGATGATGATGATGATGATGATGATGAT 360  
425 AACTGTCATTTCCCACTCCGACCTGAAATGATGATGATGATGATGATGATGATGAT 484  
361 GGTGCGCTGCTTTTGAAGGGAATTGATGATGATGATGATGATGATGATGATGATGAT 420  
485 GGTGCGCTGCTTTTGAAGGGAATTGATGATGATGATGATGATGATGATGATGATGAT 544  
421 CAAGTAGCAATATACAGAAACATGTTCAACCAATGCAAAAGTGGTGAACAGGAC 480  
545 CAAGTAGCAATATACAGAAACATGTTCAACCAATGCAAAAGTGGTGAACAGGAC 604  
481 AATGAAACAGGAATTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
605 AATGAAACAGGAATTTATATGATGATGATGATGATGATGATGATGATGATGATGAT 663  
541 GAGACATGGTTGATTCCTAGACCTGTTCCAAATTTGCTTAAGACCTTTTAAAGTTG 600

Db 664 GAGACCTGGTTGATTCCTAGACCTGTTCCAAATT--GTGTACGAGACCTTAAAGTTG 721  
QY 601 GCTGAATTTGAGCAAGATTCAAGAACATGAAACCACTATACAGAAATATTTCTTAC 660  
Db 722 GCTGAATTTGAGCAAGATTCAAGAACCTGA-----ACATATCCGAAATATTTCTCAC 775  
QY 661 AGTGAAGAACCTACTTATCTGGAATAATGAAACATCTGTTTGGGCCAACAGAAACAG 720  
Db 776 GGGGAAACCTACTCTTTTGGCGGAGTGCACA---CTGTGTTGCGGGGACCGGAAACCG 832  
QY 721 ACTCTGTTTAGCCATTAATAAGATTTT 748  
Db 833 CCCCTCGGTTCTTCTCAAGAGAAATTT 860

Search completed: April 22, 2004, 04:39:48  
Job time : 2858.76 secs



Db	661	CCGAGAAAGGGGGCAGAGACATGGTTTGATTGATTCTTACGACTGTTCCTCAAAATTTGGTTAG	720
OY	721	GACCTTTAACAAAGTTGGCTGAATTTTGGACAGAGTTCAAGAACATAGAAACCAACTATAC	780
Db	721	GACCTTTAACAAAGTTGGCTGAATTTTGGAGCAGAGTTCAAGAAACATAGAAACCAACTATAC	780
OY	781	AAGATATTTCTTTTCACTGAGGAACCTACTTATCTGGGAAATGAAACATCTGTTTTGG	840
Db	781	AAGATATTTCTTTTCACTGAGGAACCTACTTATCTGGGAAATGAAACATCTGTTTTGG	840
OY	841	GCCAAACAGGAAACAAAGACTCTTGTTGTTTACCAATAAAGATTTATTTACCCCTTCAAAAC	900
Db	841	GCCAAACAGGAAACAAAGACTCTTGTTGTTTACCAATAAAGATTTATTTACCCCTTCAAAAC	900
OY	901	ACATTTGGCAACTAAAGATTTCTGTGTAAGTCTCTTGCAAAATTTTGTATGAGATGTTGT	960
Db	901	ACATTTGGCAACTAAAGATTTCTGTGTAAGTCTCTTGCAAAATTTTGTATGAGATGTTGT	960
OY	961	GCACAAACAGTTCATTTGTTTATTTATATTTTGAATATTTGGTTTTTACTATGAATAATCCC	1020
Db	961	GCACAAACAGTTCATTTGTTTATTTATATTTTGAATATTTGGTTTTTACTATGAATAATCCC	1020
OY	1021	TTTTTATTAATATCATATGAGAAATCCCTTTACCTATCAGAAACAAAACATCTCTGG	1080
Db	1021	TTTTTATTAATATCATATGAGAAATCCCTTTACCTATCAGAAACAAAACATCTCTGG	1080
OY	1081	TTTTA TTTT	
Db	1081	TTTTA TTTT	

```

RESULT 3
US-10-653-595-26
Sequence 26, Application US/10653595
Publication No. US2004004804A1
GENERAL INFORMATION:
APPLICANT: Ruben et. al.
TITLE OF INVENTION: 95 Human secreted proteins
FILE REFERENCE: P2027P1C1
CURRENT APPLICATION NUMBER: US/10/653,595
CURRENT FILING DATE: 2003-09-03
PRIORITY APPLICATION NUMBER: US 09/397945
PRIORITY FILING DATE: 1999-09-17
PRIORITY APPLICATION NUMBER: PCT/US99/05804
PRIORITY FILING DATE: 1999-03-18
PRIORITY APPLICATION NUMBER: 60/078,566
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,576
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,573
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,574
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/078,579
PRIORITY FILING DATE: 1998-03-19
PRIORITY APPLICATION NUMBER: 60/080,314
PRIORITY FILING DATE: 1998-04-01
PRIORITY APPLICATION NUMBER: 60/080,312
PRIORITY FILING DATE: 1998-04-01
PRIORITY APPLICATION NUMBER: 60/078,578
PRIORITY FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: Patencin Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:

```

```

? NAME/KEY: SITE
? LOCATION: (1557)
? OTHER INFORMATION: n equals a,t,g, or c
?
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (1689)
? OTHER INFORMATION: n equals a,t,g, or c
?
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (1729)
? OTHER INFORMATION: n equals a,t,g, or c
?
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (1735)
? OTHER INFORMATION: n equals a,t,g, or c
?
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (1741)
? OTHER INFORMATION: n equals a,t,g, or c
?
US-10-653 -595-26

```

Query Match	98.4%	Score 1067;	DB 13;	Length 1751;
Best Local Similarity	99.6%	Pred. No. 3.2e-278;		
Matches 1067; Conservative	2;	Mismatches	2;	Indels 0; Gaps 0;

QY		GGGAGAGAGGTAAACAACGCACAGGGCGCCGAATGGGGCGGGGCGGGGCGGGAGTCGG	73
Db	16	GCGAGGAGGTAAACAACGCACAGGGCGCCGAATGGGGCGGGGCGGGGCGGGAGTCGG	75
QY	74	GAAGCGCCTTCTTGTCGTGGCGCTGGCGCTGTGGCTCGCGGTGGTTCCGGGCTGG	133
Db	76	GAAGCGCCTTCTTGTCGTGGCGCTGGCGCTGTGGCTCGCGGTGGTTCCGGGCTGG	135
QY	134	TCCCGGGTCTCGGGCATCCCCGCCGCACTGGCGGGGACCCTTAAGAAGCGTTTGAC	193
Db	136	TCCCGGGTCTCGGGCATCCCCGCCGCACTGGCGGGGACCCTTAAGAAGCGTTTGAC	195
QY	194	TTCCGTCAAACCTGATCCTTAATTGTCAAGCTAAGATACTTCTGTCCAACCTGGCTCA	253
Db	196	TTCCGTCAAACCTGATCCTTAATTGTCAAGCTAAGATACTTCTGTCCAACCTGGCTCA	255
QY	254	CCATCCCAGTTTANAGAGGGGTGANGATGATGAAGTTTTTGATTAACAAGCCCCAGTA	313
Db	256	CCATCCCAGTTTANAGAGGGGTGANGATGATGAAGTTTTTGATTAACAAGCCCCAGTA	315
QY	314	TGGGAATTTAAATGAGACCTCTGGGACACTTGAATTAATGACATGATGCCATTGGA	373
Db	316	TGGGAATTTAAATGAGACCTCTGGGACACTTGAATTAATGATGATGCCATTGGA	375
QY	374	TTGAGAAGTACATTAACCTGGCAAGAACACATGAGATGTATGAACCTTTTCCAACTT	433
Db	376	TTGAGAAGTACATTAACCTGGCAAGAACACATGAGATGTATGAACCTTTTCCAACTT	435
QY	434	GGCAACTGTACATTTCCCCATCTCCGACCTGAATGGAATCCCTTCTGTGTAAATCAA	493
Db	436	GGCAACTGTACATTTCCCCATCTCCGACCTGAATGGAATCCCTTCTGTGTAAATCAA	495
QY	494	GCGCGTCCCTCTTTTTTGAAGGAATGTATGTTCACTGGAAGAAAAATGGACATTA	553
Db	496	GCGCGTCCCTCTTTTTTGAAGGAATGTATGTTCACTGGAAGAAAAATGGACATTA	555
QY	554	GTTCAAGTGAACAATATCACGAACATGTTCAACCAATGGCAAAAGGGGTGAACACG	613
Db	556	GTTCAAGTGAACAATATCACGAACATGTTCAACCAATGGCAAAAGGGGTGAACACG	615
QY	614	GACATGAAAACAGGAATTTATTAATGACATGAAATGTAAAGCCAGCCAGAAAAGGG	673
Db	616	GACATGAAAACAGGAATTTATTAATGACATGAAATGTAAAGCCAGCCAGAAAAGGG	675
QY	674	GCAGAACATGTTTGATTCCTACGACTGTGCCAAATTTGTGTTAAGACCTTTAACAG	733
Db	676	GCAGAACATGTTTGATTCCTACGACTGTGCCAAATTTGTGTTAAGACCTTTAACAG	735



Db 796 TACAGTGGAGAACCTTACTTATCTGGGGAATGAAACATCTGTTTTGGGCGCAAGAAAC 855  
QY 854 AAGACTCTGGTTAGGCGCATAAAAAGATTATTAACCCCTTCAACCAATTTGCCAACT 913  
Db 856 AAGACTCTGGTTAGGCGCATAAAAAGATTATTAACCCCTTCAACCAATTTGCCAACT 915  
QY 914 AAGAAATTTCTGTGAGTCTCTTGCAGAAATTTTGATGACAGTGAATTTGCACAAAGCTTC 973  
Db 916 AAGAAATTTCTGTGAGTCTCTTGCAGAAATTTTGATGACAGTGAATTTGCACAAAGCTTC 975  
QY 974 TATTTGTTTATATATTTTGAATTTGTTTATTTTACCTATGAAATTCCTTTTATTAATA 033  
Db 976 TATTTGTTTATATATTTTGAATTTGTTTATTTTACCTATGAAATTCCTTTTATTAATA 035  
QY 1034 ACATATGAGAAATCCCTTACCTATCAGAAACAAACCTCTGTGTTA 1084  
Db 1036 ACATATGAGAAATCCCTTACCTATCAGAAACAAACCTCTGTGTTA 1086

RESULT 5  
US-09-122-383-13  
; Sequence 13, Application US/09122383A  
; Patent No. US20020042093A1  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Gilbertson, Debra G.  
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
; FILE REFERENCE: 97-38  
; CURRENT APPLICATION NUMBER: US/09/122,383A  
; CURRENT FILING DATE: 1998-07-24  
; EARLIER APPLICATION NUMBER: 60/053,613  
; EARLIER FILING DATE: 1997-07-24  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 13  
; LENGTH: 1038  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Degenerate nucleotide sequence encoding zslg46  
; OTHER INFORMATION: polypeptide of SEQ ID NO:2  
; FEATURE:  
; NAME/KEY: variation  
; LOCATION: (1)...(1038)  
; OTHER INFORMATION: n is any nucleotide  
US-09-122-383-13

Query Match 71.3%; Score 772.4; DB 9; Length 1038;  
Best Local Similarity 61.7%; Pred. No. 1.7e-198;  
Matches 640; Conservative 232; Mismatches 165; Indels 0; Gaps 0;

QY 47 ATGCGCGGGGGCGCGCGCGGCTCGGGGACGCGGCTCTGCTGGTGGCGGCGCTGGCGCTG 106  
Db 1 ATGCGCGGGGGCGCGCGCGGCTCGGGGACGCGGCTCTGCTGGTGGCGGCGCTGGCGCTG 60  
QY 107 CTTTGCTGCGGGTGGTTCGCGGCTGCTCCCGGCTCTCGGGGCAATCCCGTCCCGGCGCGC 166  
Db 61 YTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 120  
QY 167 TGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 226  
Db 121 TGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 180  
QY 227 AAGTACTTCTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 286  
Db 181 AAGTACTTCTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 240  
QY 287 GAAGTTTTCATTAACAAGCCCAAGTATGGAATTAATAATGAGAGACTCCCGGAGC 346  
Db 241 GAGTATTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 300  
QY 347 TTGAAATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406

Db 301 YTTAATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
QY 407 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 466  
Db 361 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
QY 467 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 526  
Db 421 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
QY 527 GTTCACTGGAAGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 586  
Db 481 GTTCACTGGAAGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 540  
QY 587 AACCAATGGAAGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 646  
Db 541 AACCAATGGAAGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAATGGAAT 600  
QY 647 AATGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 706  
Db 601 AATGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 660  
QY 707 AATTTGTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 766  
Db 661 AATTTGTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 720  
QY 767 GAAACCACTATACAAAGATATTTCTTCAAGTGAAGAACTTATCTTATCTGGAATGAA 826  
Db 721 GAAACCACTATACAAAGATATTTCTTCAAGTGAAGAACTTATCTTATCTGGAATGAA 780  
QY 827 ACATCTGTTTGGGCGCAACAGAAACAGACTCTGTTGATGACCAATTAATAATTTAT 886  
Db 781 ACATCTGTTTGGGCGCAACAGAAACAGACTCTGTTGATGACCAATTAATAATTTAT 840  
QY 887 TACCCCTTCAACCAATTTGCACTAAGAAATTTCTGTGAGTCTGTGCAATTTT 946  
Db 841 TACCCCTTCAACCAATTTGCACTAAGAAATTTCTGTGAGTCTGTGCAATTTT 900  
QY 947 GATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006  
Db 901 GATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960  
QY 1007 CCTATGAATTTCCCTTTTATTAATAATGATGAGAAATCCCTTACTATCAGAAAC 1066  
Db 961 CCTATGAATTTCCCTTTTATTAATAATGATGAGAAATCCCTTACTATCAGAAAC 1020  
QY 1067 AAAACACTCTGTGTT 1083  
Db 1021 AAAACACTCTGTGTT 1037

RESULT 6  
US-10-010-050A-13  
; Sequence 13, Application US/10010050A  
; Publication No. US20020173624A1  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Gilbertson, Debra G.  
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
; FILE REFERENCE: 97-38C1  
; CURRENT APPLICATION NUMBER: US/10/010,050A  
; CURRENT FILING DATE: 2002-03-26  
; PRIOR APPLICATION NUMBER: US 09/122,383  
; PRIOR FILING DATE: 1998-07-24  
; PRIOR APPLICATION NUMBER: US 60/053,613  
; PRIOR FILING DATE: 1997-07-24  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 1038  
; TYPE: DNA

Query Match	71.3%;	Score 772.4;	DB 14;	Length 1038;
Best Local Similarity	61.7%;	Pred. No. 1.7e-198;		
Matches 640;	Conservative 232;	Mismatches 165;	Indels 0;	Gaps 0

[illegible]

```

RESULT 7
US-09-864-761-22592
; Sequence 22592, Application US/09864761
; Patient No. US20020048763AI
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharon G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   APPLICANT: Chen, Wenheng
; TITLE OR INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22592
; LENGTH: 494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC001226.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA. SIGNAL = 4.5

```

ADDRESSEE: Genetics Institute, Inc  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A

```

? GENERAL INFORMATION:
? APPLICANT: Penn, Sharon G.
? APPLICANT: Rank, David R.
? APPLICANT: Hanzel, David K.
? APPLICANT: Chert, Wensheng
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
? TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```

```
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 5832
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC001226.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
US-09-864-761-5832

Query Match      35.1%; Score 380.2; DB 9; Length 474;
Best Local Similarity 99.2%; Pred No.2,1e-92;
Matches 382; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 748 AGCAGAGTTCAAGACATAGAACCACTATACAAAGATATTTCTTACAGTGGAGAAC 807
DB 270 AGCAGAGTTCAAGACATAGAACCACTATACAAAGATATTTCTTACAGTGGAGAAC 329
QY 808 TACTTATCTGGGAATGAAATCTGTTTGGGCCCAACAGAAACAGACTCTTGTT 867
DB 330 TACTTATCTGGGAATGAAATCTGTTTGGGCCCAACAGAAACAGACTCTTGTT 389
QY 868 AGCCATTAAGATTTTATATACCCCTTCAACCACTTGGCCACTAAGATTTCTGTT 927
DB 390 AGCCATTAAGATTTTATTTTACCCCTTCAACCACTTGGCCACTAAGATTTCTGTT 449
QY 928 GAGTCTCTTGCAAAATTTTGGATGCA 952
DB 450 GAGTCTCTTGCAAAATTTTGGATGCA 474
```

```
RESULT 10
US-10-264-049-1859
Sequence 1859, Application US/10264049
Publication No. US20040005579a1
GENERAL INFORMATION:
APPLICANT: Biree et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PAL33P1
CURRENT APPLICATION NUMBER: US/10/264,049
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: PCT/US01/18569
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: US 60/209,467
PRIOR FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 4360
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 1859
LENGTH: 497
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (175)..(175)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (175)..(175)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (285)..(285)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (295)..(295)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (313)..(313)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (323)..(324)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (331)..(331)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (395)..(395)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (440)..(440)
OTHER INFORMATION: n equals a,t,g, or c
```

```
FEATURE:
; NAME/KEY: misc feature
; LOCATION: (458)..(458)
; OTHER INFORMATION: n equals a,l,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (464)..(464)
; OTHER INFORMATION: n equals a,l,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (478)..(478)
; OTHER INFORMATION: n equals a,l,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (483)..(483)
; OTHER INFORMATION: n equals a,l,g, or c
US-10-264-049-1859
```

Query Match 24.5%; Score 266; DB 16; Length 497;  
Best Local Similarity 99.6%; Pred. No. 2e-61; Indels 0; Gaps 0;  
Matches 266; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 818 GGAATGAAACATCTGTTTGGGCGCAAGAAAGAGCTCTGGTTAGCCATPAAA 877
Db 3 GGAATGAAACATCTGTTTGGGCGCAAGAAAGAGCTCTGGTTAGCCATPAAA 62
Qy 878 AGATTTTATTAACCCCTTCAACACATTTGGCACTPAAAGATTTCTGAGTCTTG 937
Db 63 AGATTTTATTAACCCCTTCAACACATTTGGCACTPAAAGATTTCTGAGTCTTG 122
Qy 938 CAAATTTTATGATGAGTATTTGCAACACAGTTCTATTTGTTATATTTTGAATY 997
Db 123 CAAATTTTATGATGAGTATTTGCAACACAGTTCTATTTGTTATATTTTGAATY 182
Qy 998 TGGTTTATTAACCTATGAATTCCTTTTATTAATAATATGAGAAATCCCTTACCT 1057
Db 183 TGGTTTATTAACCTATGAATTCCTTTTATTAATAATATGAGAAATCCCTTACCT 242
Qy 1058 ATCAGAAACAAACACTCTCTGCTTGA 1084
Db 243 ATCAGAAACAAACACTCTCTGCTTGA 269
```

RESULT 11  
US-09-867-701-2579  
; Sequence 2579, Application US/09867701  
; Patent No. US20020132237A1  
; GENERAL INFORMATION:  
; APPLICANT: Agilite, Paul A.  
; APPLICANT: Jones, Robert  
; APPLICANT: Harlocker, Susan L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.497  
; CURRENT APPLICATION NUMBER: US/09/867,701  
; CURRENT FILING DATE: 2001-05-29  
; NUMBER OF SEQ ID NOS: 10912  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2579  
; LENGTH: 351  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-09-867-701-2579

Query Match 18.2%; Score 197.2; DB 9; Length 351;  
Best Local Similarity 84.2%; Pred. No. 7.2e-43;  
Matches 234; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

```
Qy 378 GAAGTACATTAACTGGCAAGAACTACACATGGAATGAGTGAACCTTTTCCAACTGGCA 437
Db 1 GAAGTACATTAACTGGCAAGAACTACACATGGAATGAGTGAACCTTTTCCAACTGGCA 60
Qy 438 ACTGATACATTCCCATCTCCGACCTGAATGATGATGCCCTTCTGTGTATCATCAAGCG 497
```

```
Db 61 ACTGATACATTCCCATCTCCGACCTGAATGATGATGCCCTTCTGTGTATCATCAAGCG 120
Qy 498 CTGCTGCTTTTGGAGGAATTTGATGATGTTCTACTGGAAGAAAATGGACATTAGTTC 557
Db 121 CTG-CTGCTTTTGGAGGAATTTGATGATGTTCTACTGGAAGAAAATGGACATTAGTTC 179
Qy 558 AAGTAGCACTATATACAGAAACATGTTTACCCAAATGGCAAGTGGTGAACAGAGCA 617
Db 180 AAGTAGCACTATATACAGGAATGAGTGAATATATGCAATATTTATCATCTGCATCAAA 239
Qy 618 ATGAACAGAAATTTATATGACATGATGATGATAA 655
Db 240 AACCAATGAAGAAATTTATATCTTCATTCATGAAACA 277
```

RESULT 12  
US-09-954-456-1363/c  
; Sequence 1363, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cance  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1363  
; LENGTH: 473  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-1363

Query Match 12.3%; Score 133; DB 9; Length 473;  
Best Local Similarity 100.0%; Pred. No. 2.3e-25;  
Matches 133; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 952 AGGATTTGCAACAAACAGTTCTAATTTGTTATATTTGAATATGTTGTTTACCTAT 1011
Db 473 AGGATTTGCAACAAACAGTTCTAATTTGTTATATTTGAATATGTTGTTTACCTAT 414
Qy 1012 GAAATTCCTTTTATTAATAATATGATGAGAAATCCCTTTACTATGAAACAAAC 1071
Db 413 GAAATTCCTTTTATTAATAATATGATGAGAAATCCCTTTACTATGAAACAAAC 354
Qy 1072 ACTCTCTGTTTGA 1084
Db 353 ACTCTCTGTTTGA 341
```

RESULT 13



```
US-09-908-975-10339
; Sequence 10339, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10339
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-10339

Query Match
Best Local Similarity 5.5%; Score 60; DB 10; Length 60;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 438 ACTGATCATTTCCCATCTCCGACCTGAAATGATGCCCCCTTTGTGTATATCAAGCGG 497
Db 1 ACTGATCATTTCCCATCTCCGACCTGAAATGATGCCCCCTTTGTGTATATCAAGCGG 60

RESULT 14
US-09-938-842A-5232
; Sequence 5232, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5232
; LENGTH: 1945
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-5232

Query Match
Best Local Similarity 4.2%; Score 45.4; DB 9; Length 1945;
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Cy 760 GAACATGAACCAACTATACAGAAATATTTCTTTACAGTGAGAACTTACTTATCTGGG 819
Db 121 GAATATGATTTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGATG 180

Cy 820 AAATGAACATCTGTTTGGGCGCAAGAAACAAGCTTGTGTTAGCCATTAAG 879
Db 181 AAATGATTTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGATG 180

Cy 181 AAATGATTTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGATG 879
Db 181 AAATGATTTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGATG 240
```

```
Cy 880 ATTTATTAACCCCTTGAACACATATTTGCCAATGAATTTCTGTGAGCTCTTGA 939
Db 241 GTTATTAACCCCTTGAACACATATTTGCCAATGAATTTCTGTGAGCTCTTGA 300

Cy 940 AATTTTGAATGAGATTTGCAACAACAGTTCTATTTGTTTAAATTTGAATTTG 999
Db 301 AATTTTGAATGAGATTTGCAACAACAGTTCTATTTGTTTAAATTTGAATTTG 360

Cy 1000 GTTTTAA 1006
Db 361 GCCTGTA 367

RESULT 15
US-09-938-842A-5232
; Sequence 5232, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5232
; LENGTH: 1945
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-5232

Query Match
Best Local Similarity 4.2%; Score 45.4; DB 11; Length 1945;
Matches 121; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Cy 760 GAACATGAACCAACTATACAGAAATATTTCTTTACAGTGAGAACTTACTTATCTGGG 819
Db 121 GAATATGATTTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGATG 180

Cy 820 AAATGAACATCTGTTTGGGCGCAAGAAACAAGCTTGTGTTAGCCATTAAG 879
Db 181 AAATGATTTTCACTCCCTGATATATGTTATATGAAATGAGTTTACAAAGATG 240

Cy 880 ATTTATTAACCCCTTGAACACATATTTGCCAATGAATTTCTGTGAGCTCTTGA 939
Db 241 GTTATTAACCCCTTGAACACATATTTGCCAATGAATTTCTGTGAGCTCTTGA 300

Cy 940 AATTTTGAATGAGATTTGCAACAACAGTTCTATTTGTTTAAATTTGAATTTG 999
Db 301 AATTTTGAATGAGATTTGCAACAACAGTTCTATTTGTTTAAATTTGAATTTG 360

Cy 1000 GTTTTAA 1006
Db 361 GCCTGTA 367
```

Search completed: April 22, 2004, 05:09:05  
Job time : 848.604 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 22, 2004, 00:17:57 ; Search time 92.2694 Seconds

(without alignments)  
6518.267 Million cell updates/sec

Title: US-10-010-050A-1\_COPY\_1\_1084

Perfect score: 1084  
Sequence: 1 gaattcgccagcagggcagcag.....acaacacactctctgttca 1084

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	50.6	4.7	7218	1 US-08-232-463-14	Sequence 14, Appl
C 2	41.2	3.8	640681	4 US-09-790-988-1	Sequence 1, Appl
C 3	40.6	3.7	2722	2 US-08-500-857A-7	Sequence 7, Appl
C 4	40	3.7	3877	2 US-08-599-895-1	Sequence 1, Appl
C 5	40	3.7	3877	3 US-09-211-290-1	Sequence 1, Appl
C 6	40	3.7	3877	3 US-09-322-676-1	Sequence 1, Appl
C 7	40	3.7	3877	3 US-09-466-036A-1	Sequence 1, Appl
C 8	39.6	3.7	2428	2 US-08-849-536A-3	Sequence 3, Appl
C 9	39.2	3.6	832	4 US-09-621-976-2813	Sequence 2813, Ap
C 10	39.2	3.6	8607	4 US-10-204-708-71	Sequence 71, Appl
C 11	39.2	3.6	21721	4 US-09-269-939A-41	Sequence 41, Appl
C 12	39.2	3.6	22976	4 US-09-269-939A-19	Sequence 19, Appl
C 13	39.2	3.6	23187	4 US-09-459-522-1	Sequence 1, Appl
C 14	39	3.6	429	4 US-09-252-991A-10838	Sequence 10838, A
C 15	39	3.6	474	4 US-09-252-991A-18033	Sequence 18033, A
C 16	38.2	3.5	786431	4 US-09-751-389-3	Sequence 3, Appl
C 17	37.8	3.5	543	4 US-08-956-171E-907	Sequence 907, App
C 18	37.6	3.5	4257	2 US-08-680-473-1	Sequence 1, Appl
C 19	37.6	3.5	4257	3 US-09-259-821A-1	Sequence 1, Appl
C 20	37.6	3.5	4257	3 US-08-843-659-1	Sequence 1, Appl
C 21	37.6	3.5	12001	1 US-08-458-568A-11	Sequence 11, Appl
C 22	37.6	3.5	4403765	3 US-09-103-840A-2	Sequence 2, Appl
C 23	37.2	3.4	441529	3 US-08-615-773B-4	Sequence 4, Appl
C 24	37.2	3.4	2268	4 US-08-545-528D-1	Sequence 1, Appl
C 25	36.8	3.4	580073	4 US-09-103-840A-2	Sequence 2, Appl
C 26	36.8	3.4	4403765	3 US-09-103-840A-2	Sequence 2, Appl
C 27	36.8	3.4	4403765	3 US-09-103-840A-2	Sequence 2, Appl

C 28	36.8	3.4	4411529	3 US-09-103-840A-1	Sequence 1, Appl
C 29	36.6	3.4	1830121	4 US-09-557-884-1	Sequence 1, Appl
C 30	36.6	3.4	1830121	4 US-09-643-990A-1	Sequence 1, Appl
C 31	36.4	3.4	1767	4 US-09-016-434-1468	Sequence 1468, Ap
C 32	36.4	3.4	1664976	4 US-08-916-421B-1	Sequence 1, Appl
C 33	36.2	3.3	9347	4 US-10-204-708-36	Sequence 36, Appl
C 34	36	3.3	168	3 US-08-897-924A-1	Sequence 1, Appl
C 35	36	3.3	364	4 US-09-621-976-17202	Sequence 17202, A
C 36	36	3.3	6757	4 US-09-023-655-1087	Sequence 1087, Ap
C 37	36	3.3	13086	4 US-08-956-171E-16	Sequence 16, Appl
C 38	36	3.3	43280	2 US-08-804-227C-1	Sequence 1, Appl
C 39	35.6	3.3	363	4 US-09-601-198-169	Sequence 169, App
C 40	35.4	3.3	447	4 US-09-319-014B-1	Sequence 1, Appl
C 41	35.4	3.3	832	4 US-09-621-976-2813	Sequence 2813, Ap
C 42	35.4	3.3	1164	1 US-08-416-478A-5	Sequence 5, Appl
C 43	35.4	3.3	1164	2 US-08-474-988B-5	Sequence 5, Appl
C 44	35.4	3.3	1164	2 US-08-394-442B-5	Sequence 5, Appl
C 45	35.4	3.3	1654	4 US-09-620-312D-581	Sequence 581, App

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F1s  
US-08-232-463-14  
Query Match 4.7%; Score 50.6; DB 1; Length 7218;

[illegible][illegible]

Db 698 AGCTGTGATTACAGAAAGAAATTCAAGATTCTCCAGAACAGTGTGAAGTCTGT 757  
QY 826 AACATCTGTTTGGGCCCAACAGAAACAGACTCTGTTTACCA 872  
Db 758 TTCAGTTGTTTCTATGTTTATGTTATATATGAGTATGATTCAGCTA 804

RESULT 4  
US-08-599-895-1  
; Sequence 1, Application US/08599895  
; Patent No. 5891855  
; GENERAL INFORMATION:  
; APPLICANT: Florkiewicz, Robert Z.  
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; City: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/599,895  
; FILING DATE: 31-JAN-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5891855endburg Ph.D., Carol  
; REGISTRATION NUMBER: 39,317  
; REFERENCE/DOCKET NUMBER: 760100.416  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3877 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-599-895-1

Query Match 3.7%; Score 40; DB 2; Length 3877;  
Best Local Similarity 55.9%; Pred. No. 0.37;  
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGGAGCGGCTTCTGTGCTGGGCGCTGGCGCTTTGGC 113  
Db 221 GGGGGGCGGGGCGGGGCGGCGCTGCCCCGAGCGGGTCCGAGCGCGGGCGCGGGG 280

QY 114 TCGCGGTGTTTCGGGGCTGTCCCGGGGTCTCGGGCATCCCTCCCGCGGCACTGGCCGG 173  
Db 281 ACGGCGGCTCCCGCGGCGCTCCAGCGGCTTCGGGATCCCGCGGCCCGCGAGACCA 340

QY 174 TGCCTTACAAGCGCTT 189  
Db 341 TGGCAGCGGGAGCAT 356

RESULT 5  
US-09-211-290-1  
; Sequence 1, Application US/09211290  
; Patent No. 6071885  
; GENERAL INFORMATION:  
; APPLICANT: Florkiewicz, Robert Z.  
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue  
; City: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/211,290  
; FILING DATE: 12-DEC-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Naki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 200124.401D1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3877 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-211-290-1

Query Match 3.7%; Score 40; DB 3; Length 3877;  
Best Local Similarity 55.9%; Pred. No. 0.37;  
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGGAGCGGCTTCTGTGCTGGGCGCTGGCGCTTTGGC 113  
Db 221 GGGGGGCGGGGCGGGGCGGCGCTGCCCCGAGCGGGTCCGAGCGCGGGCGCGGGG 280

QY 114 TCGCGGTGTTTCGGGGCTGTCCCGGGGTCTCGGGCATCCCTCCCGCGGCACTGGCCGG 173  
Db 281 ACGGCGGCTCCCGCGGCGCTCCAGCGGCTTCGGGATCCCGCGGCCCGCGAGACCA 340

QY 174 TGCCTTACAAGCGCTT 189  
Db 341 TGGCAGCGGGAGCAT 356

RESULT 6  
US-09-322-676-1  
; Sequence 1, Application US/09322676  
; Patent No. 6107283  
; GENERAL INFORMATION:  
; APPLICANT: Florkiewicz, Robert Z.  
; TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; City: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/322,676  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/211,290  
; FILING DATE: 12-DEC-1998

ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 200124.401D1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3877 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-322-676-1

Query Match 3.7%; Score 40; DB 3; Length 3877;  
Best Local Similarity 55.9%; Pred. No. 0.37; Mismatches 60; Indels 0; Gaps 0;  
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGAGCGCGCTTCGTGCTGCGCGCTGCGCGCTGCTTGGC 113  
DB 221 GGGGGGCGGGGCGGGGCGGCGCTGCCCCGAGCGGCTCGAGGCGGGGCGGGGCGG 280  
QY 114 TCGGGGTGTTCCGGGCTGCTCCCGGCTTCGCGGCACTCCCTCCCGCGGCACTGGCCGG 173  
DB 281 ACGGCGGCTCCCGCGGCTTCAGCGGCTCGGGGATCCCGGCGGCGCGGCA 340  
QY 174 TGCCCTACAGCGGCTT 189  
DB 341 TGGCAGCGGAGCAT 356

## RESULT 7

US-09-466-036A-1

Sequence 1, Application US/09466036A

Patent No. 6281197

GENERAL INFORMATION:

APPLICANT: Florkiewicz, Robert Z.  
TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/466,036A

FILING DATE: 17-Dec-2001

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/211,290

FILING DATE: &lt;Unknown&gt;

ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 200124.401D1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3877 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-466-036A-1

Query Match 3.7%; Score 40; DB 3; Length 3877;  
Best Local Similarity 55.9%; Pred. No. 0.37; Mismatches 60; Indels 0; Gaps 0;  
Matches 76; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 54 GGGGCGGGGCGGCGCTCGGGAGCGCGCTTCGTGCTGCGCGCTGCGCGCTGCTTGGC 113  
DB 221 GGGGGGCGGGGCGGGGCGGCGCTGCCCCGAGCGGCTCGAGGCGGGGCGGGGCGG 280  
QY 114 TCGGGGTGTTCCGGGCTGCTCCCGGCTTCGCGGCACTCCCTCCCGCGGCACTGGCCGG 173  
DB 281 ACGGCGGCTCCCGCGGCTTCAGCGGCTCGGGGATCCCGGCGGCGGCA 340  
QY 174 TGCCCTACAGCGGCTT 189  
DB 341 TGGCAGCGGAGCAT 356

## RESULT 8

US-08-849-536A-3

Sequence 3, Application US/08849536A

Patent No. 5853976

GENERAL INFORMATION:

APPLICANT: HESSE, Friederike

APPLICANT: AMEROSTUS, Dorothee

TITLE OF INVENTION: RECOMBINANT PROTEINASE FROM CLOSTRIDIUM

TITLE OF INVENTION: HISTOLYTICUM AND ITS USE FOR ISOLATING CELLS AND GROUPS OF CELLS

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nikaido, Marmelstein, Murray &amp; Oram LLP

STREET: 655 15th St., N.W., Suite 330 - G St. Lobby

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-5701

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/849,536A

FILING DATE: Herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mong, King L.

REGISTRATION NUMBER: 37,500

REFERENCE/DOCKET NUMBER: 1614-7026

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 638 - 5000

TELEFAX: (202) 638 - 4810

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 2428 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: sig\_peptide

LOCATION: 970..1026

FEATURE:

NAME/KEY: mat\_peptide

LOCATION: 1027..1965

US-08-849-536A-3

Query Match 3.7%; Score 39.6; DB 2; Length 2428;  
Best Local Similarity 51.7%; Pred. No. 0.37; Mismatches 84; Indels 0; Gaps 0;  
Matches 90; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 879 GATTATATACCCCTTCAACACATTTGCCAATAAGATTTCTGTGAGTCTTGGC 938

```
Db 1818 GAATATTTAACTCCCAAGACATTTGAAGATGCTAGAAATGCGATGATCTGCG 1877
Qy 939 AATTTTGTGACGATGATGTCACAAACAGTTCTATTGTTTAAATTTGAATAT 998
Db 1878 AATAGATTACATGCGAGAAATAGTAAACATTAACCTGCTATTAATAATCTGGGCGA 1937
Qy 999 GGTTTTACCTAATCCCTTTTATTAATAATATATAGAAATCCCTT 1052
Db 1938 TGTAGAGTTGAGAAAGAGCAGTAAATATATAGAAATATGAAGATTCATT 1991
```

## RESULT 9

```
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas MLine Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813
```

```
Query Match 3.6%; Score 39.2; DB 4; Length 832;
Best Local Similarity 9.4%; Pred. No. 0.25;
Matches 33; Conservative 174; Mismatches 142; Indels 3; Gaps 1;
```

```
Qy 584 TTCAACCAATGCGAAAGTGGTGAACGACGACATGAACGAGATTATTATGAGACA 643
Db 367 TTTGTTTWTMTWTMTTWTWTWTMTTWTMTTWTMTTWTMTTWTMTTWTMTTWTMTT 308
Qy 644 TCGAATGTAAAGCCAGCCGAGAAAGGGGCGAGACATGGTGTGATCTTACGACGTG 703
Db 307 WKBMMSTRVYCYCMKCKRRCAMTYMARGRMYSYMGKMSMSMCTMYTK 248
Qy 704 TCCAAATTTGTGTAGAAGCTTTTAAAGAGTTGGCTGAATTTGAGACAGATTCAAGAC 763
Db 247 GSTYWTMKTCACTWCYWKYKWRMSKTCMSGSRGVYTSYTSRYSMYWAMWYTC 188
Qy 764 AARGAAACCACTATACAGAAATATTTCTTACATGAGAACTACTATCTGGGAAT 823
Db 187 WWMGSMWS--TYWYMAWGKWRATWRRAMMMWMAWMTWYMMWMAWMCSSGAAAY 131
Qy 824 GAAACATGTTTGGGCGACAGAAACAAGACTCTGTTTGCCATAAAGATTT 883
Db 130 RRTMMWGRYMKRYSKRTKAMAYAKTRSYTCRMKMKCOMMMMAAAYGKT 71
Qy 884 TATTAACCCCTTCAACCACTTTGGCACTAAGAAATTTCTGTAGTCTCT 935
Db 70 MRAACWKTYYRMMWMAWMMWMTMMWYVYWRAMKRRMMWRKMSWSMMW 19
```

## RESULT 10

```
US-10-204-708-71/c
; Sequence 71, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
```

```
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 71
; LENGTH: 8607
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-71
```

```
Query Match 3.6%; Score 39.2; DB 4; Length 8607;
Best Local Similarity 48.2%; Pred. No. 1;
Matches 110; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
```

```
Qy 851 AACAGACTCTGGTTAGCCATAAAGATTATTAACCCCTTAACACCACTTGCA 910
Db 4610 ATCAAACTCTAAATTAATTAATTTCTTTCTTTTAAATAATCTTTTTC 4551
Qy 911 ACTAAGAAATCTGTGAGTCTCTGCAAAATTTTGATGACGATGTGACAAACG 970
Db 4550 CTTTAAATTTCTATTTTATTTTATTTTATTTTATTTTATTTTCTGACACT 4491
Qy 971 TTTATTTGTTTATTAATTTGAATATTTGTTTATTTTACCTATGAAATCCCTTTATTA 1030
Db 4490 TATAATTTTCTACCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 4431
Qy 1031 ATAATATGAAAGAAATCCCTTTACCTATACAGAAACAAACACTCTCT 1078
Db 4430 AATATAATACAAATAATCACTTAATCTCTTAAATAATCACTAATTTCT 4383
```

## RESULT 11

```
US-09-269-939A-41/c
; Sequence 41, Application US/09269939A
; Patent No. 6635431
; GENERAL INFORMATION:
; APPLICANT: Bihain, Bernard
; APPLICANT: Bouquellet, Lydie
; APPLICANT: Yen-Potin, Frances
; TITLE OF INVENTION: USR Receptor, Its Activity, Its Cloning and Its Application To
; TITLE OF INVENTION: The Diagnosis Prevention And/or Treatment of Obesity and
; FILE REFERENCE: GENSET.035APC
; CURRENT APPLICATION NUMBER: US/09/269,939A
; PRIOR FILING DATE: 1999-05-28
; PRIOR APPLICATION NUMBER: FR 97/10088
; PRIOR FILING DATE: 1997-08-06
; PRIOR APPLICATION NUMBER: FR 98/05032
; PRIOR FILING DATE: 1998-04-22
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent.pm
; SEQ ID NO 41
; LENGTH: 21721
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1898..2253
; NAME/KEY: exon
; LOCATION: 3438..3782
```

```

NAME/KEY: exon
LOCATION: 12064..12183
NAME/KEY: exon
LOCATION: 15049..15105
NAME/KEY: exon
LOCATION: 15670..15816
NAME/KEY: exon
LOCATION: 19486..19659
NAME/KEY: exon
LOCATION: 19806..19865
NAME/KEY: exon
LOCATION: 19963..20094
NAME/KEY: exon
LOCATION: 20236..20864
NAME/KEY: exon
LOCATION: 20954..21094
NAME/KEY: Misc_Feature
LOCATION: 715
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AC002128
NAME/KEY: Misc_Feature
LOCATION: 1229
OTHER INFORMATION: diverging insertion, G in ref genbank:AC002128
NAME/KEY: Misc_Feature
LOCATION: 3676
OTHER INFORMATION: diverging nucleotide, T in ref genbank:AC002128
NAME/KEY: Misc_Feature
LOCATION: 5039
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 5118
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 7337
OTHER INFORMATION: diverging deletion, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 8294
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 8604
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 8928
OTHER INFORMATION: diverging nucleotide, A in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 9021
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 9851
OTHER INFORMATION: diverging insertion, GATGAAA in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 9878
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 11475
OTHER INFORMATION: diverging nucleotide, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 11577
OTHER INFORMATION: diverging deletion, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 11775
OTHER INFORMATION: diverging nucleotide, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 13411
OTHER INFORMATION: diverging deletion, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 13538
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 13896
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 14912
OTHER INFORMATION: diverging nucleotide, A in ref genbank:AD000684

```

```

NAME/KEY: Misc_Feature
LOCATION: 16732
OTHER INFORMATION: diverging nucleotide, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 17169
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 18946
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 19474
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 20500
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 20501
OTHER INFORMATION: diverging deletion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 20502
OTHER INFORMATION: diverging deletion, A in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21270
OTHER INFORMATION: diverging nucleotide, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21356
OTHER INFORMATION: diverging insertion, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21476
OTHER INFORMATION: diverging nucleotide, A in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21588
OTHER INFORMATION: diverging insertion, C in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21601
OTHER INFORMATION: diverging deletion, T in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 21635
OTHER INFORMATION: diverging insertion, G in ref genbank:AD000684
NAME/KEY: Misc_Feature
LOCATION: 19963..19965
OTHER INFORMATION: Potential variant splicing site AAC
NAME/KEY: Misc_Feature
LOCATION: 1..21721
OTHER INFORMATION: n = a,g,c or t
US-09-269-939A-41

```

```

Query Match 3.6%; Score 39.2; DB 4; Length 21721;
Best Local Similarity 54.1%; Pred. No. 1.7;
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

```

Qy 31 GGCACAGGCGCCGAGATGCGCGCGCGCGCGCTTGGGAGCGCGCTTCTGCTG 90
Db 21327 GGGCGGGGCGCAGTGGCAGCGCCCTGGGCGGGCGGCGACGACGCGCGCTGG 21268
Qy 91 CTGGCGCTGGCGCTGCTTGGCTCGCGGCGGCTTCCGGGCTGCTCCGGGCTGGGCGAT 150
Db 21267 CTGGCGCGGAGACGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 21208
Qy 151 CCCCTCCGCGCGACGCTGCGCGGCTGCC 178
Db 21207 TTCTCCGCGCGCGCGACTCCGATGCC 21180

```

```

RESULT 12
US-09-269-939A-19/c
Sequence 19, Application US/09269939A
Patent No. 6635431
GENERAL INFORMATION:
APPLICANT: Bihain, Bernard
APPLICANT: Bouguetelret, Lydie
APPLICANT: Yen-Potin, Frances
TITLE OF INVENTION: LSR Receptor, Its Activity, Its Cloning and Its Application To

```



```

1 EARLIER FILING DATE: 1999-07-20
2 NUMBER OF SEQ ID NOS: 20
3 SOFTWARE: Patent.pm
4 SEQ ID NO: 1
5 LENGTH: 23187
6
7 TYPE: DNA
8 ORGANISM: Homo sapiens
9 FEATURE:
10 NAME/KEY: exon
11 LOCATION: 3540..3884
12 OTHER INFORMATION: exon2
13 FEATURE:
14 NAME/KEY: exon
15 LOCATION: 12163..12282
16 OTHER INFORMATION: exon3
17 FEATURE:
18 NAME/KEY: exon
19 LOCATION: 15144..15200
20 OTHER INFORMATION: exon4
21 FEATURE:
22 NAME/KEY: exon
23 LOCATION: 15765..15911
24 OTHER INFORMATION: exon5
25 FEATURE:
26 NAME/KEY: exon
27 LOCATION: 19579..19752
28 OTHER INFORMATION: exon6
29 FEATURE:
30 NAME/KEY: exon
31 LOCATION: 19899..19958
32 OTHER INFORMATION: exon7
33 FEATURE:
34 NAME/KEY: exon
35 LOCATION: 20056..20187
36 OTHER INFORMATION: exon8
37 FEATURE:
38 NAME/KEY: exon
39 LOCATION: 20329..20957
40 OTHER INFORMATION: exon9
41 FEATURE:
42 NAME/KEY: exon
43 LOCATION: 21047..21187
44 OTHER INFORMATION: exon10
45 FEATURE:
46 NAME/KEY: polyA signal
47 LOCATION: 21168..21173
48 OTHER INFORMATION: AATAAA
49 FEATURE:
50 NAME/KEY: misc feature
51 LOCATION: 1..2000
52 OTHER INFORMATION: potential 5'regulatory region
53 FEATURE:
54 NAME/KEY: misc feature
55 LOCATION: 22324..23187
56 OTHER INFORMATION: homology with USF2 gene in ref: emb1 Y07661
57 FEATURE:
58 NAME/KEY: primer bind
59 LOCATION: 523..544
60 OTHER INFORMATION: upstream amplification primer 17-2
61 FEATURE:
62 NAME/KEY: primer bind
63 LOCATION: 1047..1168
64 OTHER INFORMATION: downstream amplification primer 17-2, complement
65 FEATURE:
66 NAME/KEY: primer bind
67 LOCATION: 946..963
68 OTHER INFORMATION: upstream amplification primer 99-4576
69 FEATURE:
70 NAME/KEY: primer bind

```

LOCATION: 1385..1402  
OTHER INFORMATION: downstream amplification primer 99-4576 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 1096..1115  
OTHER INFORMATION: upstream amplification primer 9-19  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 1616..1635  
OTHER INFORMATION: downstream amplification primer 9-19 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 1602..1621  
OTHER INFORMATION: upstream amplification primer 9-20  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 2074..2093  
OTHER INFORMATION: downstream amplification primer 9-20 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 2036..2053  
OTHER INFORMATION: upstream amplification primer 99-4557  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 2563..2580  
OTHER INFORMATION: downstream amplification primer 99-4557 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 2084..2102  
OTHER INFORMATION: upstream amplification primer 9-1  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 2483..2500  
OTHER INFORMATION: downstream amplification primer 9-1 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 2470..2489  
OTHER INFORMATION: upstream amplification primer 9-21 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 2062..2081  
OTHER INFORMATION: downstream amplification primer 9-21  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 3455..3474  
OTHER INFORMATION: upstream amplification primer 9-3  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 3882..3901  
OTHER INFORMATION: downstream amplification primer 9-3 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 3775..3792  
OTHER INFORMATION: upstream amplification primer 99-4558  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 4336..4356  
OTHER INFORMATION: downstream amplification primer 99-4558 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 4902..4920  
OTHER INFORMATION: upstream amplification primer 99-14419 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 4444..4463  
OTHER INFORMATION: downstream amplification primer 99-14419  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 6638..6655  
OTHER INFORMATION: upstream amplification primer 99-4577  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 7072..7089

OTHER INFORMATION: downstream amplification primer 99-4577 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 7995..8012  
OTHER INFORMATION: upstream amplification primer 99-4559  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 8576..8593  
OTHER INFORMATION: downstream amplification primer 99-4559 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 9622..9639  
OTHER INFORMATION: upstream amplification primer 99-3148  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 10023..10040  
OTHER INFORMATION: downstream amplification primer 99-3148 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 9964..9981  
OTHER INFORMATION: upstream amplification primer 99-4560  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 10996..11015  
OTHER INFORMATION: upstream amplification primer 99-14411 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 10546..10563  
OTHER INFORMATION: downstream amplification primer 99-4560 , complement  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 10492..10512  
OTHER INFORMATION: downstream amplification primer 99-14411  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 11972..11990  
OTHER INFORMATION: upstream amplification primer 99-4561  
FEATURE:  
NAME/KEY: primer\_bind  
LOCATION: 12481..12501  
OTHER INFORMATION: downstream amplification primer 99-4561 , complement

Query Match 3.6%; Score 39.2; DB 4; Length 23187;  
Best Local Similarity 54.1%; Pred. No. 1.8;  
Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 31 GGCAACGGGCGCGAGATCGGCGGGCGCGGCGCTCGGAGACGGCTTCTGTTG 90  
DB 21420 GGCGCGGGCGCGATGCGACGCTGCGCGCGCGCGCGCGACCGCGGCGGCGTGG 21361  
QY 91 CTGGGCGTGGCGCTTGGCTCGGCGGCGTTCGCGGCTGCTCCGGGCTCGGCGAT 150  
DB 21360 CTGCGCGGCGACGTTCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTC 21301  
QY 151 CCCCTCCGCGCGCGCACTGCGCGGTCGCC 178  
DB 21300 TTCCTCCGCGCGCGCGACCTCGGTCGCC 21273

RESULT 14  
US-09-252-991A-10838/c  
; Sequence 10838, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10838  
; LENGTH: 429  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10838

Query Match 3.6%; Score 39; DB 4; Length 429;  
Best Local Similarity 51.4%; Pred. No. 0.19;  
Matches 90; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 5 TCGGCACGAGGGGACAGAGTAGACACAGGACAGAGGCGCCGAGATGCGGCGGGCGGGGC 64  
DB 314 TCGCCCGCGCTGGGCGGCGGTACGCGGAGATGTGGGTGCCGACAAAGCCGATCTCGGG 255  
QY 65 GCGGCTCGGGGACGCGCTTCTGTGTGCTGGGCGCTGCGCTGCTTTGGCTCGCGGTGTT 124  
DB 254 GCTGGGCTTTGTTGGGCGCGCTGGCGCTGCGCTCGCGCGGTGATGTTGGCGGTG 195  
QY 125 CCGGCTGTGTCGCGGCTTCGCGCATCCCTCCGCGCCACTGCGCGGATGCCCT 179  
DB 194 CGGTGATGCTGCTGGGCTTGGCGCGCGCGCTGCTGCTTCCGCGCGCT 140

RESULT 15  
US-09-252-991A-10890  
; Sequence 10890, Application US/09252991A  
; Patent No. 6551795.  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10890  
; LENGTH: 474  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10890

Query Match 3.6%; Score 39; DB 4; Length 474;  
Best Local Similarity 51.4%; Pred. No. 0.21;  
Matches 90; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 5 TCGGCACGAGGGGACAGAGTAGACACAGGACAGAGGCGCCGAGATGCGGCGGGCGGGGC 64  
DB 176 TCGCCCGCGCTGGGCGGCGGTACGCGGAGATGTGGGTGCCGACAAAGCCGATCTCGGG 235  
QY 65 GCGGCTCGGGGACGCGCTTCTGTGTGCTGGGCGCTGCGCTGCTTTGGCTCGCGGTGTT 124  
DB 236 GCTGGGCTTTGTTGGGCGCGCTGGCGCTGCGCTCGCGCGGTGATGTTGGGCGGTG 295  
QY 125 CCGGCTGTGTCGCGGCTTCGCGCATCCCTCCGCGCCACTGCGCGGATGCCCT 179  
DB 296 CGGTGATGCTGCTGGGCTTGGCGCGCGCGCTGCTGCTTCCGCGCGCT 350

Search completed: April 22, 2004, 04:42:56  
Job time : 110.289 secs



CC the thyroid. This product can be used to study secretion of proteins from  
 CC cells and also to treat or prevent deficient expression of zsig46, which  
 CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'  
 CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that  
 CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's  
 CC disease). Anticardiolipin antibodies, Wilson disease and Reiger  
 CC syndrome). Antibodies and other binding proteins, are used as immunoassay  
 CC reagents to detect zsig46 or cells expressing it, e.g. for assessing  
 CC thyroid function to produce anti-idiotypic antibodies, for affinity  
 CC purification of zsig46, to screen expression libraries, to neutralise  
 CC zsig46 activity, and to deliver toxins, radioisotopes etc. for  
 CC therapeutic or diagnostic purposes. Agonists of the product can be used  
 CC to promote growth, differentiation and proliferation of specific cell  
 CC types, e.g. for treating (extra)thyroid diseases or as additive to cell  
 CC cultures  
 CC  
 XX Sequence 1486 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;

Query Match 100.0%; Score 1084; DB 2; Length 1486;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-271;  
 Matches 1084; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTGGGACGAGGGGCGAGAGTACACGCGACAGGGCGCCGAGATGCGCGGCGCGC 60  
 DB 1 GAATTGGGACGAGGGGCGAGAGTACACGCGACAGGGCGCCGAGATGCGCGGCGCGC 60  
 QY 61 GGGCGCGGCTCGGGGACGCGCTTCTGGTGGCGGGGCGCGCTTGGCTCGCGGT 120  
 DB 61 GGGCGCGGCTCGGGGACGCGCTTCTGGTGGCGGGGCGCGCTTGGCTCGCGGT 120  
 QY 121 GGTTCGCGGCTGGTCCGCGGCTTCGCGGACCTCCCTCCGCGGCGCACTGGCGGCTCCCTA 180  
 DB 121 GGTTCGCGGCTGGTCCGCGGCTTCGCGGACCTCCCTCCGCGGCGCACTGGCGGCTCCCTA 180  
 QY 121 GGTTCGCGGCTGGTCCGCGGCTTCGCGGACCTCCCTCCGCGGCGCACTGGCGGCTCCCTA 180  
 DB 121 GGTTCGCGGCTGGTCCGCGGCTTCGCGGACCTCCCTCCGCGGCGCACTGGCGGCTCCCTA 180  
 QY 181 CAAAGCGCTTTGACTTCGCTCCAAAACCTGATCTTATTTGTAAGCTAATGTAATCTTCTG 240  
 DB 181 CAAAGCGCTTTGACTTCGCTCCAAAACCTGATCTTATTTGTAAGCTAATGTAATCTTCTG 240  
 QY 241 TCCAACTGGCTCACCTATCCGATGAGAGGGGATGATGACATGTAAGTTTTCGATT 300  
 DB 241 TCCAACTGGCTCACCTATCCGATGAGAGGGGATGATGACATGTAAGTTTTCGATT 300  
 QY 301 ACAAGCCCCAGTATGGAATTTAATATGAGACCTCTCGGAGCACTTGAAATTATGCA 360  
 DB 301 ACAAGCCCCAGTATGGAATTTAATATGAGACCTCTCGGAGCACTTGAAATTATGCA 360  
 QY 361 TGAATCCATTGGATTCAAGATGATTAAGTGGCAAGACTACCAATGGAATGGTATGA 420  
 DB 361 TGAATCCATTGGATTCAAGATGATTAAGTGGCAAGACTACCAATGGAATGGTATGA 420  
 QY 421 ACTTTTCAACCTGGCACTGATCACTTCCCATCTCGGAGCTGAAATGATGCCCTT 480  
 DB 421 ACTTTTCAACCTGGCACTGATCACTTCCCATCTCGGAGCTGAAATGATGCCCTT 480  
 QY 481 CTGGTGTATCAAGCGCTGCTGCTTTTGAAGGAATTGATGATTTCTCGTGAAGA 540  
 DB 481 CTGGTGTATCAAGCGCTGCTGCTTTTGAAGGAATTGATGATTTCTCGTGAAGA 540  
 QY 541 AAATGGGACATTAGTTCAAGTACCACTATACAGGAACATGTTCAACAAAGGCGAAA 600  
 DB 541 AAATGGGACATTAGTTCAAGTACCACTATACAGGAACATGTTCAACAAAGGCGAAA 600  
 QY 601 GTGGGTGAACAGGACATGAAACAGAAATTTATATGACATGGAATGTAAGCCAG 660  
 DB 601 GTGGGTGAACAGGACATGAAACAGAAATTTATATGACATGGAATGTAAGCCAG 660  
 QY 661 CCCGAGAAAAGGGGAGAGACATGGTTGATTCCTACACATGTTCCAAATTTGTATAG 720  
 DB 661 CCCGAGAAAAGGGGAGAGACATGGTTGATTCCTACACATGTTCCAAATTTGTATAG 720  
 QY 721 GACCTTTAAACAAGTTGGTGAATTTGGAGAGATTGAAGAACTATGAATAC 780  
 DB 721 GACCTTTAAACAAGTTGGTGAATTTGGAGAGATTGAAGAACTATGAATAC 780

QY 781 AAGATATTTCTTACAGTGGAGAACCTATCTATCTGGGAAATGAACATCTGTTTTGG 840  
 DB 781 AAGATATTTCTTACAGTGGAGAACCTATCTATCTGGGAAATGAACATCTGTTTTGG 840  
 QY 841 GCCACAGAAAACAAGACTCTGTTTACGATTAAGCAATTAATTAATCCCTCAAAAC 900  
 DB 841 GCCACAGAAAACAAGACTCTGTTTACGATTAAGCAATTAATTAATCCCTCAAAAC 900  
 QY 901 ACATTTGCCACATTAAGAAATTTCTGTTGAGTCTTCCGAATTTTGTATGACGATTTGT 960  
 DB 901 ACATTTGCCACATTAAGAAATTTCTGTTGAGTCTTCCGAATTTTGTATGACGATTTGT 960  
 QY 961 GCAAAACAGTCTAATTTGTTTATTAATTTGAATATTTGTTTACCTATGAATGCC 1020  
 DB 961 GCAAAACAGTCTAATTTGTTTATTAATTTGAATATTTGTTTACCTATGAATGCC 1020  
 QY 1021 TTTTATTAATAATTAATATGAGAAATCCCTTACCTATGAGAAACAAACACTCTCTG 1080  
 DB 1021 TTTTATTAATAATTAATATGAGAAATCCCTTACCTATGAGAAACAAACACTCTCTG 1080  
 QY 1081 TTTTA 1084  
 DB 1081 TTTTA 1084

## RESULT 2

AA224825  
 ID AA224826 standard; DNA, 1751 BP.

AA224826;

02-DEC-1999 (first entry)

Human secreted protein gene 16 clone HMZADJ7.

Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

MO9947540-A1.

23-SEP-1999.

18-MAR-1999; 99MO-US005804.

19-MAR-1998; 98US-0078563P.  
 19-MAR-1998; 98US-0078566P.  
 19-MAR-1998; 98US-0078573P.  
 19-MAR-1998; 98US-0078574P.  
 19-MAR-1998; 98US-0078576P.  
 19-MAR-1998; 98US-0078577P.  
 19-MAR-1998; 98US-0078578P.  
 19-MAR-1998; 98US-0078579P.  
 19-MAR-1998; 98US-0078581P.  
 01-APR-1998; 98US-0080312P.  
 01-APR-1998; 98US-0080313P.  
 01-APR-1998; 98US-0080314P.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;  
 Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;  
 Shi Y, Moore PA.  
 WPI; 1999-562050/47.

NR P-PSDB; AAY41323.  
XX New isolated human genes, useful for diagnosis and treatment of e.g.  
PT cancers, neurological disorders, immune diseases, inflammation or blood  
PT disorders.  
XX  
XX Claim 1; Page 308; 484pp; English.

XX This sequence represents a nucleic acid molecule which encodes a secreted  
XX human protein. The gene number, and the clone it is derived from, are  
XX detailed in the descriptor line. The gene can be used to generate fusion  
XX proteins by linking to the gene to a human immunoglobulin Fc portion  
XX (e.g. AA224802) for increasing the stability of the fused protein as  
XX compared to the human protein only. The invention relates to 95 novel  
XX genes and their fragments (nucleic acid sequences: AA224811-224907; amino  
XX acid sequences AA14308-Y41404) which are useful for preventing, treating  
XX or ameliorating medical conditions e.g. by protein or gene therapy. Also,  
XX pathological conditions can be diagnosed by determining the amount of the  
XX new polypeptides in a sample or by determining the presence of mutations  
XX in the new polynucleotides. Specific uses are described for each of the  
XX 95 polynucleotides, based on which tissues they are most highly expressed  
XX in (see AA224811 for described uses)

XX Sequence 1751 BP; 477 A; 355 C; 380 G; 523 T; 0 U; 16 Other;

XX Query Match 98.4%; Score 1067; DB 2; Length 1751;

XX Best Local Similarity 99.6%; Pred. No. 7.9e-267; Mismatches 2; Indels 0; Gaps 0;

XX Matches 1067; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
2Y 14 GGGGAGAGGTAGACACGCGCCGAGATGCGCGCGCGCGCGCTCG 73
2b 16 GCGGAGAGGTAGACACGCGCCGAGATGCGCGCGCGCGCGCTCG 75
2Y 74 GAGCGCGCTTCTGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTG 133
2b 76 GAGCGCGCTTCTGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTG 135
2Y 134 TCCGCGGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCTGCTGAC 193
2b 136 TCCGCGGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCTGCTGAC 195
2Y 194 TTCCGTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 253
2b 196 TTCCGTCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 255
2Y 254 CCTATCCGAGTTATGAGGCGGTATGATGACATGGAATTTTCATTA 313
2b 256 CCTATCCGAGTTATGAGGCGGTATGATGACATGGAATTTTCATTA 315
2Y 314 TGGGAAATTTAAATGAGAGCTCTGCGGACACTTGAATTTATGATGCA 373
2b 316 TGGGAAATTTAAATGAGAGCTCTGCGGACACTTGAATTTATGATGCA 375
2Y 374 TTCAAGAGTACATTAATGAGAGCAATGATGATGATGATGATGATGAT 433
2b 376 TTCAAGAGTACATTAATGAGAGCAATGATGATGATGATGATGATGAT 435
2Y 434 GCGCACTGTAATTTCCCATCTCCGACCTGAAATGATGCCCTTTCTGT 493
2b 436 GCGCACTGTAATTTCCCATCTCCGACCTGAAATGATGCCCTTTCTGT 495
2Y 494 GCGCGTGCCTGCTTTTGAAGGAAATGATGATGATGATGATGATGAT 533
2b 496 GCGCGTGCCTGCTTTTGAAGGAAATGATGATGATGATGATGATGAT 535
2Y 554 GTTCAAGTACATTAATGAGAGCAATGATGATGATGATGATGATGAT 613
2b 556 GTTCAAGTACATTAATGAGAGCAATGATGATGATGATGATGATGAT 615
2Y 614 GACATGAGAGCAATTTATTTATGAGAGCAATGATGATGATGATGAT 673
2b 616 GACATGAGAGCAATTTATTTATGAGAGCAATGATGATGATGATGAT 675
```

```
QY 674 GCAGAGACATGTTGATTCCTACGACTGTTCCAAATTTGTTAAGGACCTTAA 733
Db 676 GCAAGACATGTTGATTCCTACGACTGTTCCAAATTTGTTAAGGACCTTAA 735
QY 734 TTGCTGTAATTTGAGAGCAAGTTCAAGAACATAGAACCAATACAGATATT 793
Db 736 TTGCTGTAATTTGAGAGCAAGTTCAAGAACATAGAACCAATACAGATATT 795
QY 794 TACGATGAGAGCAATCTTATCTGAGAAATGAAACATCTGTTTGGGCCA 853
Db 796 TACGATGAGAGCAATCTTATCTGAGAAATGAAACATCTGTTTGGGCCA 855
QY 854 AAGACTCTGTTAGCCATTAATAAAGATTTATATACCCCTCAACACATTTG 913
Db 856 AAGACTCTGTTAGCCATTAATAAAGATTTATATACCCCTCAACACATTTG 915
QY 914 AAGAAATTTCTGTTAGCTCTTCCAAATTTTATGAGAGTATTTGACAA 973
Db 916 AAGAAATTTCTGTTAGCTCTTCCAAATTTTATGAGAGTATTTGACAA 975
QY 974 TATTTGTTTATATTTTGAATATGTTTATTTTACCTATGAAATTCCTT 1033
Db 976 TATTTGTTTATATTTTGAATATGTTTATTTTACCTATGAAATTCCTT 1035
QY 1034 ACATATGAGAGAAATCCCTTATCTATCAGAAACAAACACTCTGTTTA 1084
Db 1036 ACATATGAGAGAAATCCCTTATCTATCAGAAACAAACACTCTGTTTA 1086
```

#### RESULT 3

AAK94829  
ID AAK94829 standard; cDNA; 2120 BP.

AAK94829;

06-NOV-2001 (first entry)

Human full-length cDNA, SEQ ID NO: 3977.

Human, full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EP1130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-00114089.

08-JUL-1999; 99JP-00194486.

11-JAN-2000; 2000JP-00118774.

02-MAY-2000; 2000JP-00183765.

(HELI-) HELIX RES INST.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

WPI; 2001-524255/58.

P-PSDB; AAM93870.

830 Primers useful for synthesizing full length cDNA clones and their use

in genetic manipulation.

Claim 8; SEQ ID NO 3977; 1380bp + Sequence Listing; English.

The invention relates to primers for synthesizing full length cDNA clones, 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesized by the oligo-capping method. The primers



CC enable the production of the full length cDNA easily without any special  
CC methods. The present sequence is a full length human cDNA of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in CD-ROM format directly  
CC from EPO

XX Sequence 2120 BP; 590 A; 409 C; 460 G; 661 T; 0 U; 0 Other;

Query Match 98.2%; Score 1064.6; DB 4; Length 2120;  
Best Local Similarity 99.6%; Pred. No. 3.6e-266;  
Matches 1067; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 14 GGGGAGAGGATGACACAGGCGCCGAGATGCGGCGGCGGCGGCGGCGGCTCGG 73
DB 19 GCGGAGAGGATGACACAGGCGCCGAGATGCGGCGGCGGCGGCGGCGGCTCGG 78
QY 74 GAGCGCGCTCTGCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 133
DB 79 GAGCGCGCTCTGCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 138
QY 134 TCCCGGAGTCTGCGGCGATCCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCTTGG 193
DB 139 TCCCGGAGTCTGCGGCGATCCCTCCCGGCGGCGGCGGCGGCGGCGGCGGCTTGG 198
QY 194 TTCCGTCGAAACCTGATCCTTATTTGCAAGCTAAGTACTTTCTTCTGCACTGCTCA 253
DB 199 TTCCGTCGAAACCTGATCCTTATTTGCAAGCTAAGTACTTTCTTCTGCACTGCTCA 258
QY 254 CCTATCCAGTATGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATG 313
DB 259 CCTATCCAGTATGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATG 318
QY 314 TGGGAAATTAATAATGAGGACCTCTCTGGAACCTTGAATAATGATGATGATGATG 373
DB 319 TGGGAAATTAATAATGAGGACCTCTCTGGAACCTTGAATAATGATGATGATGATG 378
QY 374 TTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 433
DB 379 TTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 438
QY 434 GGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 493
DB 439 GGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 498
QY 494 GGGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 553
DB 499 GGGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 558
QY 554 GTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 613
DB 559 GTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 618
QY 614 GACATGAAACAGGATTTTATGAGACATGAGATGAAAGCCAGCCAGAAAGGGG 673
DB 619 GACATGAAACAGGATTTTATGAGACATGAGATGAAAGCCAGCCAGAAAGGGG 678
QY 674 GCGAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 733
DB 679 GCGAGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 738
QY 734 TTGCGTGAATTTGAGAGAGGATTCAGAAACATGAGAAACCAATGAGAAATTTTCT 793
DB 739 TTGCGTGAATTTGAGAGAGGATTCAGAAACATGAGAAACCAATGAGAAATTTTCT 798
QY 794 TACAGTGAAGACCTTATGAGAGAGGATTCAGAAACATGAGAAACCAATGAGAA 853
DB 799 TACAGTGAAGACCTTATGAGAGAGGATTCAGAAACATGAGAAACCAATGAGAA 858
QY 854 AAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 913
DB 859 AAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 918
QY 914 AAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 973

```

```

DB 919 AAGAAATTTCTGTTGAGTCTCTGCAAAATTTTGTGATGATGATGATGATGATG 978
QY 974 TATTTGTTTATATATTTTGAATATGATGATGATGATGATGATGATGATGATGAT 1033
DB 979 TATTTGTTTATATATTTTGAATATGATGATGATGATGATGATGATGATGATGAT 1038
QY 1034 ACATATGAAGAAATCCCTTACCTATGAGAAACAAACACCTCTGCTTGA 1084
DB 1039 ACATATGAAGAAATCCCTTACCTATGAGAAACAAACACCTCTGCTTGA 1089

```

#### RESULT 4

AA02866  
ID AA02866 standard; DNA; 1038 BP.

AC AA02866;

DT 14-MAY-1999 (first entry)

DE Human degenerate zsig46 DNA.

KM Secreted protein; zsig46; human; chromosome 13; thyroid disease;  
KM hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;  
KM Hirschsprung's disease; neuronal ceroid-lipofucinosi; Wilson disease;  
KM Reiger syndrome; immunoassay; detection; anti-idiotypic antibody;  
KM therapy; diagnostic; ss.

OS Homo sapiens.

PN MO9905275-A1.

PD 04-FEB-1999.

PF 24-JUL-1998; 98WO-US015431.

PR 24-JUL-1997; 97US-0053613P.

PA (ZYMO) ZYMOGENETICS INC.

PI Sheppard PO, Gilbertson DG;

DR WPI; 1999-142930/12.

PT New secreted polypeptide, zsig46, and its fragments, related fusion  
PT proteins - used for diagnosis and treatment of thyroid disorders or  
PT diseases involving genes on chromosome 13.

PS Claim 31; Page 94-95; 101bp; English.

CC This invention describes the isolation of a novel human secreted protein,  
CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in  
CC the thyroid. This product can be used to study secretion of proteins from  
CC cells and also to treat or prevent deficient expression of zsig46, which  
CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'  
CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that  
CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's  
CC disease, neuronal ceroid-lipofucinosi, Wilson disease and Reiger  
CC syndrome). Antibodies and other binding proteins, are used as immunoassay  
CC reagents to detect zsig46 or cells expressing it, e.g. for assessing  
CC thyroid function to produce anti-idiotypic antibodies, for affinity  
CC purification of zsig46, to screen expression libraries, for neutralise  
CC zsig46 activity, and to deliver toxins, radioisotopes etc. for  
CC therapeutic or diagnostic purposes. Agonists of the product can be used  
CC to promote growth, differentiation and proliferation of specific cell  
CC types, e.g. for treating (extra)thyroid diseases or as additive to cell  
CC cultures

XX Sequence 1038 BP; 189 A; 106 C; 179 G; 166 T; 0 U; 398 Other;

Query Match 71.3%; Score 772.4; DB 2; Length 1038;  
Best Local Similarity 61.7%; Pred. No. 2.4e-190;  
Matches 640; Conservative 232; Mismatches 165; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - nucleic search, using frame\_plus\_p2n model  
on on: April 24, 2004, 22:59:32 ; Search time 6540.36 Seconds  
(without alignments)  
2292.945 Million cell updates/sec

Title: US-10-010-050A-2  
Effect score: 1927  
Sequence: 1 MRGAGAGRGRASWCNALAL.....IKTYEIPPIPRNKTLSGL 346

Coring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 2167151695 residues  
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Out-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
MODEL=frame+ p2n.model -DEV=xlh  
Q=/cgm2.1/USPTO.spool/US10010050/runat\_22042004\_113204\_27549/app.query.fasta\_1.1372  
DB=GenBank1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCUT=0 -LOOPEXT=0  
UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.cdt -LIST=45  
DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
OUTFMT=pro -NORF=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
USBR=US10010050\_@CCN\_1\_1\_7433\_@runat\_22042004\_113204\_27549 -NCPU=6 -ICPU=3  
NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: GenBank1:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_pi:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_to:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pin:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_higo\_hum:\*  
40: em\_higo\_mus:\*  
41: em\_higo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1927	100.0	1486	6 BD073402	BD073402 Secreted
2	1927	100.0	4080	9 AF068227	AF068227 Homo sapi
3	1916	99.4	1751	6 BD136338	BD136338 95 human
4	1913	99.3	2120	6 BD127858	BD127858 Primer fo
5	1913	99.3	2120	9 AK075109	AK075109 Homo sapi
6	1631	84.6	1038	6 BD073413	BD073413 Secreted
7	1451	75.3	2318	10 BC025487	BC025487 Mus muscu
8	1425	73.9	198524	2 AC109554	AC109554 Rattus no
9	1368	71.0	227920	2 AC107510	AC107510 Rattus no
10	1065	55.3	697	6 BD125187	BD125187 Primer fo
11	1065	55.3	697	6 BD126453	BD126453 primer fo
12	925.5	48.0	169362	9 AC001226	AC001226 Genomic s
13	925.5	48.0	251187	2 AL136440	AL136440 Homo sapi
14	925.5	48.0	251187	2 AL135987	AL135987 Homo sapi
15	730	37.9	110000	10 AB014175	AB014175 Mus muscu
16	730	37.9	156823	2 AC102815	AC102815 Mus muscu
17	727	37.7	506	6 BD060465	BD060465 Secreted
18	716	37.2	229015	2 AC106966	AC106966 Rattus no
19	711	36.9	217336	2 AC131344	AC131344 Rattus no
20	617.5	32.0	251187	2 AL135987	AL135987 Homo sapi
21	582	30.2	198524	2 AC109554	AC109554 Rattus no
22	272.5	14.1	258273	2 AC113855	AC113855 Rattus no
23	256	13.3	110000	3 AC116305	AC116305 Dictyoste
24	241	12.5	473	6 AX333881	AX333881 Sequence
25	241	12.5	473	6 AX335993	AX335993 Sequence
26	146.5	7.6	256879	3 AC116582	AC116582 Dictyoste
27	117	6.1	112027	2 AC007006	AC007006 Homo sapi
28	115	6.0	187727	2 AC064806	AC064806 Homo sapi
29	112	5.8	2162	2 BC014298	BC014298 Homo sapi
30	112	5.8	2175	6 AX881019	AX881019 Sequence
31	112	5.8	2175	6 BD158729	BD158729 Primer fo
32	112	5.8	2175	6 AK023964	AK023964 Homo sapi
33	111.5	5.8	124548	8 AP004083	AP004083 Oryza sat
34	111.5	5.8	223182	5 AL292904	AL292904 Zebrafish
35	110	5.7	236195	2 AC073713	AC073713 Mus muscu
36	109	5.7	145500	3 LMFL4803	LMFL4803
37	109	5.7	145500	2 AC137990	AC137990 Leishmani
38	108.5	5.6	263129	2 AC098289	AC098289 Rattus no
39	108	5.6	204545	2 AC108566	AC108566 Rattus no
40	106	5.5	110000	2 AL160016	AL160016 Continuation (4 of
41	106	5.5	130244	9 AL590705	AL590705 Human DNA
42	105	5.4	155156	9 AF241727	AF241727 Homo sapi
43	104.5	5.4	143847	2 AF025589	AF025589 Homo sapi
44	104.5	5.4	191078	2 AC022505	AC022505 Homo sapi
45	104	5.4	2061	9 HSM805115	AL833837 Homo sapi

## ALIGNMENTS

RESULT 1

	LOCUS	BD073402	1486 bp	DNA	linear	PAT 27-AUG-2002
	DEFINITION	Secreted protein which human chromosome 13 encodes.				
	ACCESSION	BD073402.1	GI:22619005			
	VERSION	JP 2001511345-A/1.				
	KEYWORDS	Homo sapiens (human)				
	SOURCE	Homo sapiens				
	ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
	REFERENCE	Shppard,P.O. and Gilberton,D.G. Patent : JP 2001511345-A 1 14-AUG-2001;				
	AUTHORS	ZYMOGENETICS INC				
	TITLE	JZYMOMERICS INC				
	JOURNAL	OS Homo sapiens (human) PN JP 2001511345-A/1 PD 14-AUG-2001 PF 24-JUL-1998 JP 2000504249 PI PAUL O SHEPPARD,DIBRA G GILBERTON PC C12N15/09,A61K38/00,A61K48/00,C07KL4/47,C07KL6/18,C12N1/15, C12N1/19,				
	FEATURES	source FT CDS Location/Qualifiers 1..1486 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"				
	ORIGIN					
	Alignment Scores:					
	Pred. No.:	1.41e-166 Length: 1486				
	Score:	1927.00 Matches: 346				
	Percent Similarity:	100.00% Conservative: 0				
	Best Local Similarity:	100.00% Mismatches: 0				
	Query Match:	100.00% Indels: 0				
	DB:	Gaps: 0				
	US-10-010-050A-2 (1-346) x BD073402 (1-1486)					
Qy	1	MetArGArGAlYaIagIyaLaalAArGlyARgAlaseRTpCYSTPalaLeuAlaLau	20			
Db	47	ATTGGGGGGGGCGGGGGGGGGCCGTGGGAAGGCCTTCCTGCTCTGGGGCTGGGGCTG	106			
Qy	21	LLeuTRpleuaLaValvalProGLYTTrnSerArgValserGLyllepRoserARGARHis	40			
Db	107	CTTTGGGCTCGGGTGGTGTCGGGGTCCGGGGTCCGGGCCAACCCCCTCCGGCGCAC	166			
Qy	41	TriPRoVALPrOTryLYLSarGPheASPheargPRoylSProaspPROtyrCYSGlnALA	60			
Db	167	TGGCGGGGCCCCtTCAAGCCGCTTGCATCTCCGTCCAACAACCtGAtTctTAATGcAAGCT	226			
Qy	61	LyfTyThrPhCySPROThngLYsePRollEPRoVALMetGlugLYasPASpaSIle	80			
Db	227	AAGTATACTTCTCTCCAACACTGGCTCACTATCCAGTATAGGAGGGTGATGATGACATT	286			
Qy	81	GlutALpheARGueLengInAlaPRoVALTrpgLuBhelSYTYrgLYAspLeuleNGLYHis	100			
Db	287	GAACTTTTTGCATTAACAAGCCCAAGTATGGGAATTTAAAATAAGAAGAcCTTCGGAAC	346			
Qy	101	IeuLYsileMeTHisAPalaIllegLYphEARgeSRThrLeuthRGilyLASNTyrThr	120			
Db	347	TTGAATAATTAATGCATGATGACCATGATGATGAGAGTAGATTAACCTGGCAAGAACTAACA	406			
Qy	121	MeGLUthrPYrgLUleuPheGInLeuGLYasnCYstHRPhePROHIslEuarPGlu	140			
Db	407	ATGGAATAGATAGACTTTTCCAACCTGGCAACTGTACATTTCCTCCCAATCTCCGACTGAA	466			

Oy		141	MelarsinA1arProPheTyrCysAsnGlnGlyAlaIalaCysPhePheGluGlyIleAsp	160
Db		467	ATGGATGCCCTTTCTGTGTAAATCAAGGGCTCCCTGCTTTTTTGAGGAATTGATGAT	526
Oy		161	VahistripLySGluAenGlyThrIeuValGlnValAlaThriIeSerGlyAsnMetPhe	180
Db		527	GTTCACTGGAAGAAGAAATGGACACTTAGTTCAAGTAGCACTATATCATGGAACAATGTC	586
Oy		181	AsnGlnMetAlaLysTrpValLysGlnAspAsnGlnPhcGlyIleTyrrGluInThrTrp	200
Db		587	AACCAAATGGCAAAAGTGGTGAACACAGACAATGAACAGAAATTTATTAAGACACATGG	646
Oy		201	AsnValIylsAlaserProGluLysGlyAlaGluThrTrpPheaSerpTyrAspCysSer	220
Db		647	AATATAAAGCGACCCGCAGAAAAGGGGCGAGAGCATGTTGATTCCTACGATGCTTCC	706
Oy		221	LysPheValIeuArgThrPheaAsnLysLeuAlaGluDheGlyAlaGluPheLysAsnIle	240
Db		707	AAATTTGTGTTAAGACCTTTAACAAAGTGGCTGAATTTGGACAGAGATTCAGAACATA	766
Oy		241	GluThrAsnLyrThrArgGliePheLeuTyrSerGlyGluProMhrTyrlleuGlyAsnGlu	260
Db		767	GAACCCAACATACMAAAATATTTCTTTACACGGGAGAACCTTATCTGGGGAATGAA	826
Oy		261	ThrservalPheglYProthrGlyAsnLysThrIeuGlyLeuAlaIleLysArgPheTyr	280
Db		827	ACATCTGTTTTGGGGCCACAGGAACAAACACTCTTGTTAGCCAAAAAAGATTTAT	886
Oy		281	TyrTrpPheLysProHisLeuProThrLysGluPheLeuIeuSerLeuLeuGlnIlePhe	300
Db		887	TACCCTTCAAAACCACTTGGCAACTAAAGAAATTTGTTGAGTCTCTTGCAAAATTTT	946
Oy		301	AspAlaValIleValHisLysGlnPheTyrlleuPheTyrAsnPheGluTyrTrpPheLyu	320
Db		947	GATGCACTGATGTGGCACAAACAGTCTATTTGTTATATATTTGAAATATGGTTTTTA	1006
Oy		321	ProMetLysPheProPheIleLysIleThrTyrgluGluIleProLeuProIleArgAsn	340
Db		1007	CCATGGAATTCCTCTTTATTAATAATACATATGAGAAATTCCTTTACCTATCATGAAC	1066
Oy		341	LysThrIeuSerGlyLeu 346	
Db		1067	AAAACACTCTCTGGTTTA 1084	
RESULT 2				
AF068227				
LOCUS				
DEFINITION				
DESCRIPTION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCES				
AUTHORS				
TITLE				
JOURNAL				











y	201	AasnallyslaaeerProgluysglyalagluThrTTPheAspSerTyAAspcySer	226
b	601	AAATGTAARCGMNSCCNGARRARGSGCNGARACCTGGTTTGATGWSNTAGATGYMSN	660
y	221	LysPheValIleuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle	240
y	661	AAATTGTGTYTMGNMCANTTTAAVYAARYTMGCNGARTTGGGCGCNGARTTAAABAAYATH	720
y	241	GluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu	260
y	721	GARCAANAATAACMGNAATHTTTTTYNTATYWSNGGARCCNACTAATYATWNGNAAGAR	780
b	261	ThServalPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr	280
b	781	ACNNMSGNTMTTGGGCCACCGNAATAAACACTYTMGNYTMGCNATHAARMGNTTAY	840
y	281	TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePhe	300
b	841	TATCCNTTYYAARCCNCATYTMCCNACNARAGRTTYYTWTWSYNTNTNCARATHTTY	900
y	301	AspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu	320
y	901	GAYCGCNGTNATHGTCNCAVYARCARTTYATYATYTATYAAATTTGARARAATGGTTTNTN	960
y	321	ProMetLysPheProPheIleLysIleThrTyrGluGlnIleProLeuProIleArgsn	340
b	961	CCTATGAARTTCCNTTYYATHAAPATHACNTATGAGARGARTRCCNTYMCNATHMGNAAY	1020
y	341	LysThrLeuSerGly 345	
b	1021	AAACAATYTMNSNGCN 1035	
ESULT 7	BC025487	2318 bp mRNA linear ROD 16-APR-2003	
OCUS	EFINITION	Mus musculus cereoid-lipofuscinosis, neuronal 5, mRNA (cDNA clone IMAGE:5251891), partial cds.	
CESSION	BRSION	BC025487	
EYMORDS	OURCE	BC025487.1 GI:19343790	
ORGANISM		Mus musculus (house mouse)	
REFERENCE		Mus musculus	
AUTHORS		Eudaryota; Eutelezia; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2318)	
		Strauberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shamen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhut,N.K., Hopkins,R.F., Jordan,H., Moore,T., Wax,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marisina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Caesavart,T.L., Scheerer,T.E., Brownstein,M.U., Ushid,T.B., Toshiyuki,S., Cantucci,P., Prange,C., Raha,S.S., Locunellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKeenan,K.U., Malek,U.A., Gunaratne,P.H., Richard,S., Worsley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalona,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahney,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,D.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Gittwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalski,U., Smallus,D.E., Schermer,A., Schein,J.E., Jones,S.J., and Marra,M.A.	
		Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences	
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16699-16903 (2002)	
PUBMED		22389257	
REFERENCE		12477932	
AUTHORS		2 (bases 1 to 2318)	
TITLE		Strauberg,R.	
JOURNAL		Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer	

```
REMARK
COMMENT
NHL-MGC Project URL: http://mgc.nci.nih.gov
Email: cgaaps-remail.nih.gov
Contact: MGC help desk
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Ahler, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dieterich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R.,
Maduro, O.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J.D., Pearson, R., Stathitop, S., Thomas, P.J., Touchman, J.W.,
Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNI at: http://image.llnl.gov
Series: IRAX Plate: 56 Row: C Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
location/Qualifiers
1. .2318
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5251891"
/tissue_type="Mammary tumor. MAP-TGF alpha model. 7 months
old. gross tissue."
/clone_id="NCI CGAP Mam5"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
<1. .2318
/gene="Cln5"
/db_xref="LOCUSID:211286"
/db_xref="MGI:2442253"
<1. .1013
/gene="Cln5"
/codon_start=3
/product="Cln5 protein"
/protein_id="AAH25487.1"
/db_xref="RefSeq:NM_001003"
/db_xref="MGI:2442253"
/db_xref="MGI:2442253"
/translation="TRPAHRPALALIGLATIGASPTSGORPVYKRPSPKX
DPYQAKYTCRPGSPPIPMKDNVITVLRLQAIMEKYDGLGHEFMKADNGFRP
TLTKENATTEMTFLPOLGNCTFFPLRPDKAPFNOCGAAPFBESIDMKKNSTLL
VVALISNTENTKAEMVKQNDETIIYYETVVRAGGAGATWESIDCSNFLETTY
KLAFGEFGFKIEINVTIKIPLYGSEPIYLGNETSIFGGKNNKTALAIKKYGPFRP
LSTDPLMNFLKIPDTVIIRQFYLFYNFEYMLPMKPEPVKIITYEETPLPTRHTTF
DL"
CD5
gene
CDS
ORIGIN
Alignment Scores:
Pred..No.: 8.63e-123 Length: 2318
Percent: 1451.00 Matches: 258
Score Similarity: 85.76% Conservative: 31
Best Local Similarity: 76.56% Mismatch: 42
Query Match: 75.30% Indels: 6
DB: 10 Gaps: 2
US-10-010-050A-2 (1-346) x BC025487 (1-2318)
```

```

12 GCGCACTGGCGCCGCGCTCGCGCTGGCGCTGGCGCTGGCGCAAGATTCTGGCGCG
30 SerArgValSerGlyLysLeuProSerArgArgHisSTripProValProTyrLysArgPheAsp 49
72 TCCCGCGCGCTGGCG-----CAACGCTGGCGCGCGCTCGCAAGCGCTTCTCT 113
50 PheArgProLysProAspProTyrCysGlnAlaValTyrThrPheCysProThrGlySer 69
120 TTCGGCGCGCAAGACAGATCCCTTCTCTCAAGCTAAGTATATCTTGTCTACCGCGCTCG 173
70 ProIleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProVal 89
180 CCCATCCCGATTATGAGAGCAATGACATGATCGAGGCTTTCAGACTACAAAGCCCGGATT 233
90 TrpGluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGly 103
240 TGGGAATTTAATATGAGAGCTCTCGGAGACCTTTAACTTATGATGACCGCGCGGA 293
110 PheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrGlyLeuPheGlnLeu 123
300 TTCAGAGGACACTGACAGCGCAAGACTACAAATGAGTGTATGACTTTTCCAGCTG 359
130 GlyAsnGlyThrPheProHisLeuAlaGlyProGluMetCaspAlaProPheTyrCysAsnGln 149
360 GGCAGACTGATCATTTCCCGCCTCCGCGCTGCAAGAGCGCTCTCTGCTGTAACTCAA 419
150 GlyAlaAlaCysPhePheGluGlyIleAspAspValHisSTripLysGluAsnGlyThrLeu 169
420 GGGGCGCGCTCTTTTTCAGAGAAATGATGATAAACCTCGAAGAGAAACGGGACACTG 479
170 ValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrValLysGln 189
480 TCAGTGGTTCACACATATCCGAAACACATTTAAACAAAGTGGCGCGAGGAGCGAG 539
190 AspAsnGluThrGlyIleTyrTyrGluThrThrPasnValLysAlaSerProGluLysGly 209
540 GACAAATGAACTGGGATTTATATGAGACATGACGCTCGGAGCGCGCCGCAAGCAAGG 599
210 AlGluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLys 229
600 GCCAGAGCGGATTCGAGTCTCGACACTGTTCCAAATTTTCTTAAAGACATATAGAAA 659
230 LeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeu 249
660 TTGGCTGAATTTGGAACAGAAATTCAGAGATGAAACAAACTATACGAAATATATTTCTT 719
250 TyrSerGlyGluProTyrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsn 269
720 TACAGTGGAGGCTTATTTACCTGGGAAATGAAACATCTATTTTGGGCCCAAGGAAAC 779
270 LysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThr 289
780 AAGACTCTGCTTGGCCATTAATAAATTTATAGCCCTTCAGACCGGTATTTGTCAAC 839
290 LysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPhe 309
840 AAAGATTCTTCGATGATATTTCTTCAAAATTTTATATACAGATATATACAGACACTTC 859
310 TyrLeuPheTyrAsnPheGluTyrTyrThrPheLeuProMetLysPheProPheIleLysIle 329
900 TACTTGTATTAATACCTTGAATATTTGTTCTTACCAATGAAACCCCTTGTCAAAATA 959
330 ThrTyrGluGluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 346
960 ACATACGAGAAACCCGCTTACTACCCGACATACACACTTACGACTTG 1010

```

```

RESULT 8
AC109554/ 198524 bp DNA linear HTG 11-OCT-2002
LOCUS Rattus norvegicus clone CH230-331D3, *** SEQUENCING IN PROGRESS
DEFINITION ***
AC109554
AC109554 GI:23820683
VERSION

```

KEYWORDS: HTG, HTGS, PHASE2, HTGS\_DRAFT, HTGS\_ENRICHED.

SOURCE: Rattus norvegicus (Norway rat)

ORGANISM: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE: 1 (bases 1 to 198524)  
Muzny, D., Marie, M., Metker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, M., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anha, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escobedo, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Hernandez, R., Hines, S., Hladun, S., Hodgson, N., Hernandez, J., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hognes, M., Hollins, B., Howells, S., Hui, Y., Hume, U., Idle, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowalski, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lohanshewa, L., Louisedge, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munnadas, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwackemele, O., Okunolu, G., Olarinmoye, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polidexter, A., Popovic, D., Prims, E., Pu, L., L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ritz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shutsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sibson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, U., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valae, R., Verez, V., Villaseas, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wiczyski, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, U., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

TITLE: Direct Submission

JOURNAL: 2 (bases 1 to 198524)

REFERENCE: 2 (bases 1 to 198524)  
Worley, K.C.  
Direct Submission

JOURNAL: Submitted (05-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE: 3 (bases 1 to 198524)  
Rat Genome Sequencing Consortium.  
Direct Submission

JOURNAL: Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT: On Oct 11, 2002 this sequence version replaced gi:21738217. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas





```

37360 TTCTCCCTCCGTCGAGACAGATCCCTTCTGTCAAGCCAACTATCTTCTGCCCTACT 37419
68 GlycerProileProvalMetGluGlyAspAspAspIleGluValPheArgLeuGlnAla 87
37420 GGCTACACCATCCCGATTATGAGACACATGATCATCGAAGTCTTAAGGCTACAAACC 37479
88 ProValTTPGluPheLeuTyGlyAspLeuLeuGlyHisLeuValIleMetHisAspAla 107
37480 CCGTTTGGGAATTAATAATGAGACCTCCGCGAGACCTTCAAAATCATGACAGATCC 37539
108 IleglyPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTTPYrGluLeuPhe 127
37540 ATTGGATTGAGAGATCCCTGACCGGCAAGACTACACATTAATGATGATGAGCTTTTC 37599
128 GlnLeuGlyAsnCyThrPheProHisLeuArgProGluMetAspAlaProPheTTPCyS 147
37600 CAGCTCGGCAACTGATACATTTCCCACTCGCGCGAAGGAAAGCGTCCGTTGCTGTGT 37659
148 AsnGlnGlyAlaAlaAcYpPhePheGluGlyIleAspAspValHisTTPYrGluAsnGly 167
37660 AACCGAGGGGCGAGCTCTTTGAGAGAAATTTACGATTAACACTGGAAGGAAAAACGGG 37719
168 ThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGluMetAlaIleTTPYrVal 187
37720 ACGTTGCTCTGCTGTCGACACATATCCGAAACACGTTTAACAAATGAGCGAGTGGGTG 37779
188 LysGlnAspAsnGluThrGlyIleTyrTyrGluThrTTPAsnValIleAspSerProGlu 207
37780 AAGCAGACAAATGACATCGGATTTATTAACGACATGAGACATGCGGCGCCAGCCCAAGA 37839
208 LysGlyAlaGluThrTTPheAspSerTyrAspCySerLysPheValLeuArgThrPhe 227
37840 AAGGGGGGCGAGACATGTTTGAATCTTAACGACTGTGCAATTTGTCTTAAGGCAATAT 37899
228 AsnLeuLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIle 247
37900 GAGAAATGGCTGCAATTTGGAACAGATTCAGAAAGATGAAACAAACATATATGAAATA 37959
248 PheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThr 267
37960 TTTCTTTACAGGGGAGAAACCGATTTACTTGGAAATGAAACGTCATTTTGGGCCCAA 38019
268 GlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeu 287
38020 GGAACCAAGACCTTGTGCTTGGCCATTAATAAATTTTACGGCCCGTCMAAGGATTCG 38079
288 ProThrLysGluPheLeuLeuSerLeuGlnIlePheAspAlaValIleValHisLys 307
38080 TCACCAAGAGATTTCTGTGTAATTTCTGAAATTTTGTGACACAGTGAATATGACACA 38139
308 GlnPheTyrLeuPheTyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheIle 327
38140 GAGTCTACCTGTTTATTAACCTTGGATGTGTTGTTCTACCTATGAAACCCCTTTGTC 38199
38200
38300
38400
38500
38600
38700
38800
38900
39000
39100
39200
39300
39400
39500
39600
39700
39800
39900
40000
40100
40200
40300
40400
40500
40600
40700
40800
40900
41000
41100
41200
41300
41400
41500
41600
41700
41800
41900
42000
42100
42200
42300
42400
42500
42600
42700
42800
42900
43000
43100
43200
43300
43400
43500
43600
43700
43800
43900
44000
44100
44200
44300
44400
44500
44600
44700
44800
44900
45000
45100
45200
45300
45400
45500
45600
45700
45800
45900
46000
46100
46200
46300
46400
46500
46600
46700
46800
46900
47000
47100
47200
47300
47400
47500
47600
47700
47800
47900
48000
48100
48200
48300
48400
48500
48600
48700
48800
48900
49000
49100
49200
49300
49400
49500
49600
49700
49800
49900
50000
50100
50200
50300
50400
50500
50600
50700
50800
50900
51000
51100
51200
51300
51400
51500
51600
51700
51800
51900
52000
52100
52200
52300
52400
52500
52600
52700
52800
52900
53000
53100
53200
53300
53400
53500
53600
53700
53800
53900
54000
54100
54200
54300
54400
54500
54600
54700
54800
54900
55000
55100
55200
55300
55400
55500
55600
55700
55800
55900
56000
56100
56200
56300
56400
56500
56600
56700
56800
56900
57000
57100
57200
57300
57400
57500
57600
57700
57800
57900
58000
58100
58200
58300
58400
58500
58600
58700
58800
58900
59000
59100
59200
59300
59400
59500
59600
59700
59800
59900
60000
60100
60200
60300
60400
60500
60600
60700
60800
60900
61000
61100
61200
61300
61400
61500
61600
61700
61800
61900
62000
62100
62200
62300
62400
62500
62600
62700
62800
62900
63000
63100
63200
63300
63400
63500
63600
63700
63800
63900
64000
64100
64200
64300
64400
64500
64600
64700
64800
64900
65000
65100
65200
65300
65400
65500
65600
65700
65800
65900
66000
66100
66200
66300
66400
66500
66600
66700
66800
66900
67000
67100
67200
67300
67400
67500
67600
67700
67800
67900
68000
68100
68200
68300
68400
68500
68600
68700
68800
68900
69000
69100
69200
69300
69400
69500
69600
69700
69800
69900
70000
70100
70200
70300
70400
70500
70600
70700
70800
70900
71000
71100
71200
71300
71400
71500
71600
71700
71800
71900
72000
72100
72200
72300
72400
72500
72600
72700
72800
72900
73000
73100
73200
73300
73400
73500
73600
73700
73800
73900
74000
74100
74200
74300
74400
74500
74600
74700
74800
74900
75000
75100
75200
75300
75400
75500
75600
75700
75800
75900
76000
76100
76200
76300
76400
76500
76600
76700
76800
76900
77000
77100
77200
77300
77400
77500
77600
77700
77800
77900
78000
78100
78200
78300
78400
78500
78600
78700
78800
78900
79000
79100
79200
79300
79400
79500
79600
79700
79800
79900
80000
80100
80200
80300
80400
80500
80600
80700
80800
80900
81000
81100
81200
81300
81400
81500
81600
81700
81800
81900
82000
82100
82200
82300
82400
82500
82600
82700
82800
82900
83000
83100
83200
83300
83400
83500
83600
83700
83800
83900
84000
84100
84200
84300
84400
84500
84600
84700
84800
84900
85000
85100
85200
85300
85400
85500
85600
85700
85800
85900
86000
86100
86200
86300
86400
86500
86600
86700
86800
86900
87000
87100
87200
87300
87400
87500
87600
87700
87800
87900
88000
88100
88200
88300
88400
88500
88600
88700
88800
88900
89000
89100
89200
89300
89400
89500
89600
89700
89800
89900
90000
90100
90200
90300
90400
90500
90600
90700
90800
90900
91000
91100
91200
91300
91400
91500
91600
91700
91800
91900
92000
92100
92200
92300
92400
92500
92600
92700
92800
92900
93000
93100
93200
93300
93400
93500
93600
93700
93800
93900
94000
94100
94200
94300
94400
94500
94600
94700
94800
94900
95000
95100
95200
95300
95400
95500
95600
95700
95800
95900
96000
96100
96200
96300
96400
96500
96600
96700
96800
96900
97000
97100
97200
97300
97400
97500
97600
97700
97800
97900
98000
98100
98200
98300
98400
98500
98600
98700
98800
98900
99000
99100
99200
99300
99400
99500
99600
99700
99800
99900
100000

```

```

PN JP 2002017375-A/618
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI MAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/
10, C12P1/02, C12P1/68, C12P2/08, G06F17/30, C12N15/00, C12N5/00 CC
PC Primer for synthesizing full-length cDNA and use thereof FH Key
FT source 1..697
FT location/Qualifiers
FT 1..697 /organism='Homo sapiens (human)'.
FEATURES
source location/Qualifiers
1..697 /organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Alignment Scores:
Pred. No.: 376-88 Length: 697
Score: 1065.00 Matches: 200
Percent Similarity: 93.55% Conserved: 3
Best Local Similarity: 92.17% Mismatches: 8
Query Match: 55.27% Indels: 6
DB: Gaps: 2
US-10-010-050A-2 (1-346) x BD125187 (1-697)
1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTTPYrTTPAlaLeuAlaLeu 20
52 ATGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGGCGGGGG 111
21 LeuTTPLeuAlaValAlaProGlyTTPSerArgValSerGlyIleProSerArgArgHis 40
112 CTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 171
41 TTPProValProGlyLysArgPheAspPheArgProLysProAspProTyrCySglnAla 60
172 TGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 231
61 LysTyrThrPheCySProThrGlySerProIleProValMetGluGlyAspAspAspIle 80
232 AAGTATATCTTCTGTGCACTGCTCACTATCCAGTTATGAGGGGATGATGATGACT 291
81 GluValPheArgLeuGlnAlaProValTTPGluPheLysTyrGlyAspLeuLeuGlyHis 100
292 GAGCTTTTTCATTAACAAGCCCAAGTATGGAATTTAATATGAGACCTCTGGGACAC 351
101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
352 TTGAAATATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
121 MetGluTTPYrGluLeuPheGlnLeuGlyAsnCyThrPheProHisLeuArgProGlu 140
412 ATGGAATGATTAAGACTTTTCCAACTTGGCACTGCACTTATTTCCCACTCTCGA 471
141 MetAspAlaProPheTTPCyAsnGlnIleLysAlaCyPhePheGluGlyIleAspAsp 160
472 ATGGATGGCCCTTCTGTGATATTAAGCGCTGCTCTCTTTTGGAGGAATTAATGAT 531
161 ValHisTTPYrGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetP 180
532 GTTCACTGGAAGGAAAGGAGCACTTAAGTTCAATAGCACTATATCAAGGAAACATGT 591
180 heAsnGlnMetAla-LysTTPValLysGln--AspAsnGluThrGlyIleTyrTyrGlu 198
592 TCACCAATATGCAAAAGTGGGTGAAACAGACACATTAAGAAACAAGAAATTTATATGAG 651

```







repeat_region	/rpt family="CAT"n"
	complement(20389. .20494)
repeat_region	/rpt family="FAM"
	21238. .21534
repeat_region	/rpt family="AluSx"
	complement(12331. .23353)
repeat_region	/rpt family="GC rich"
	complement(12346. .23578)
repeat_region	/rpt family="GC rich"
	2465. .24707
repeat_region	/rpt family="MIR"
	2603. .26047
repeat_region	/rpt family="MIR2"
	complement(12717. .28532)
repeat_region	/rpt family="MIRB1"
	complement(12853. .28854)
repeat_region	/rpt family="AluSg"
	complement(12860. .23076)
repeat_region	/rpt family="A1"
	complement(12965. .23561)
repeat_region	/rpt family="MIR"
	complement(30050. .30147)
repeat_region	/rpt family="MIR"
	31487. .31812
repeat_region	/rpt family="AluDo"
	complement(33629. .33732)
repeat_region	/rpt family="MIR"
	3443. .34729
repeat_region	/rpt family="AluSx"
	35070. .35361
repeat_region	/rpt family="AluSx"
	complement(36437. .36461)
repeat_region	/rpt family="AT rich"
	complement(39967. .40261)
repeat_region	/rpt family="AluSx"
	complement(40741. .40915)
repeat_region	/rpt family="AluSg"
	40963. .41663
repeat_region	/rpt family="L1PA2"
	42042. .42252
repeat_region	/rpt family="MTIE"
	42467. .42771
repeat_region	/rpt family="AluSg"
	42772. .42848
repeat_region	/rpt family="L1MB8"
	42849. .43151
repeat_region	/rpt family="AluDo"
	43152. .43318
repeat_region	/rpt family="L1MA10"
	43253. .43427
repeat_region	/rpt family="L1MB8"
	43429. .43672
repeat_region	/rpt family="L1MB8"
	complement(44188. .44481)
repeat_region	/rpt family="AluSc"
	4454. .45246
repeat_region	/rpt family="L1TR8"
	complement(44699. .47032)
repeat_region	/rpt family="AT rich"
	complement(4741. .47431)
repeat_region	/rpt family="AT rich"
	complement(47612. .47684)
repeat_region	/rpt family="CAT"n"
	complement(47946. .48239)
repeat_region	/rpt family="AluDo"
	complement(49827. .49872)
repeat_region	/rpt family="AT rich"
	complement(50160. .50186)
repeat_region	/rpt family="AT rich"
	50952. .51121
repeat_region	/rpt family="MER5A"
	complement(52140. .52284)
repeat_region	/rpt family="AluDo"

```

repeat_region      53772..53974
repeat_region      /rpt_family="MER1B"
repeat_region      54225..54348
repeat_region      /rpt_family="MIR"
repeat_region      complement(54533..54579)
repeat_region      /rpt_family="(TGAA)n"
repeat_region      complement(54580..54736)
repeat_region      /rpt_family="AluJb"
repeat_region      complement(54755..54782)
repeat_region      /rpt_family="(CA)n"

Alignment Scores:
pred. No.:          1,41e-72      Length:          106988
score:              925.50         Matches:          173
Percent Similarity: 89.69%         Conservative:     1
Best Local Similarity: 89.18%       Mismatches:      19
Query Match:        48.03%          Indels:          19
JB:                 9              Gaps:            1

JS-10-010-050a-2 (1-346) x AC001226 (1-106988)
2Y      153  CysPhePheGluGly1LeaspPheValHisTPLeuGluAsnGlyThrLeuValGlnVal 172
Db      31851 TGTGTTTTTAAA----- 31362

2Y      173  AleThrIleSerGlyAsnMetPheAsnGlnMerAlaIysTrpValIysGlnAspAsnGln 192
Db      31863 -----CTAGGAAACATGTTTCACCAATGCGAAAGGGGTGAAACAGACATGAA 31913

2Y      193  ThrGlyIleTyTrpGluThrTPAsnValIysAlaSerProGluIysGlyAlaGluThr 212
Db      31914 ACAGGAATTATATATGAGCATGGAATGTAAAGCCAGCCAGAAAGGGGCGAGAGAC 31973

2Y      213  TyrPheAspSerTyTrpAspCySerIysPheValLeuArgThrPheAsnIysLeuAlaGln 222
Db      31974 TGGTTGATTCCTTACGACCTGTTCCAAATTTGTGTAAGACCTTTAACAGTTGGCGTAA 32033

2Y      233  PheGlyAlaGluPheIysAsnIleGluThrAsnTyThrArgIlePheLeuTySerGly 252
Db      32034 TTGAGACAGAGTTCAAGACATGAAACCACTATCAAGAAATTTCTTTACAGTGA 32093

2Y      253  GluProThrTyTrpLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIysThrLeu 272
Db      32094 GAACCTATCTTACGGGAAATGAACATCTGTTTGGGCAACAGAAACAAGACTCTT 32153

2Y      273  GlyLeuAlaIleIysArgPheTyTrpProPheIysProHisLeuProThrIysGluPhe 292
Db      32154 GGTTAGCCATAAAGATTTTATTACCCCTTCAACACATTTGCCAAGCTAAAGAAATTT 32213

2Y      293  LeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisIysGlnPheTyLeuPhe 312
Db      32214 CTGTTGAGTCTCTGCAAAATTTTGAAGCAGTGTGTCGCAAAACAGTTCTATTGTTT 32273

2Y      313  TyrAsnPheGluTyTrpPheLeuProMetIysPheProPheIleIysIleThrTyGlu 332
Db      32274 TATATATTTGAAATTTGGTTTTTACCTATGAAATTCCTTTTATAAATAATCATATGAA 32333

2Y      333  GluIleProLeuProIleArgAsnIysThrLeuSerGlyLeu 346
Db      32334 GAAATCCCTTACCTATCATGAAACAAACACTCTCTGTGTTA 32375

RESULT 13
AL136440      169362 bp      DNA      linear      HTG 10-JUL-2001
LOCUS      Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.
DEFINITION      AL136440
ACCESSION      AL136440.2 GI:10039473
VERSION      HTG; HTGS PHASE1; HTGS_CANCELED.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1

```

```

AUTHORS      Burton,J.
TITLE      Direct Submission
JOURNAL      Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT      requests: clonerequest@sanger.ac.uk
On Sep 8, 2000 this sequence version replaced gi:6982057.
-----
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
Center project name: ba185124
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 162085 bases at least Q40
Consensus quality: 164521 bases at least Q30
Consensus quality: 166040 bases at least Q20
Insert size: 168062; sum-of-contigs
Insert size: 164966; 1.7% error; agarose-fp
Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality
Coverage: 4.16x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1      3400: contig of 3400 bp in length
*      3401      3500: gap of 100 bp
*      3501      16444: contig of 12944 bp in length
*      16445      16544: gap of 100 bp
*      16545      19514: contig of 2970 bp in length
*      19515      19614: gap of 100 bp
*      19615      22241: contig of 2627 bp in length
*      22242      22341: gap of 100 bp
*      22342      25526: contig of 3185 bp in length
*      25527      25626: gap of 100 bp
*      25627      33070: contig of 7444 bp in length
*      33071      33170: gap of 100 bp
*      33171      67620: contig of 34450 bp in length
*      67621      67720: gap of 100 bp
*      67721      93413: contig of 25693 bp in length
*      93414      93513: gap of 100 bp
*      93514      114081: contig of 20368 bp in length
*      114082      114181: gap of 100 bp
*      114182      126206: contig of 12025 bp in length
*      126207      126306: gap of 100 bp
*      126307      144583: contig of 18277 bp in length
*      144584      144683: gap of 100 bp
*      144684      147136: contig of 2453 bp in length
*      147137      147236: gap of 100 bp
*      147237      164925: contig of 17689 bp in length
*      164926      165025: gap of 100 bp
*      165026      169362: contig of 4337 bp in length.
*
Location/Qualifiers
1.169362
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone_lib="RP11-185124"
/clone_1lb="RPCT-11.1"
1.3400
/note="assembly_fragment:00318
clone_end:17
vector_side:left"
3501..16444
misc_feature
misc_feature
misc_feature

```

```

/misc_feature /note="assembly_fragment:00726
fragment_chain:1"
16545..19514
/misc_feature /note="assembly_fragment:00157
fragment_chain:1"
19615..22241
/misc_feature /note="assembly_fragment:00846
fragment_chain:1"
22342..25526
/misc_feature /note="assembly_fragment:00271
fragment_chain:1"
25627..33070
/misc_feature /note="assembly_fragment:01555
fragment_chain:1"
33171..67620
/misc_feature /note="assembly_fragment:00768"
67721..93413
/misc_feature /note="assembly_fragment:00828"
93514..114081
/misc_feature /note="assembly_fragment:01026"
114182..126206
/misc_feature /note="assembly_fragment:01600
fragment_chain:2"
126307..144583
/misc_feature /note="assembly_fragment:01198
fragment_chain:2"
144684..147136
/misc_feature /note="assembly_fragment:01132
fragment_chain:2"
147237..164925
/misc_feature /note="assembly_fragment:01005
fragment_chain:2"
165026..169362
/misc_feature /note="assembly_fragment:01413
fragment_chain:2"
clone_end:SP6
vector_side:right"
ORIGIN
Alignment Scores:
red. No.: 2.55e-72 Length: 169362
core: 925.50 Matches: 173
Percent Similarity: 89.69% Conservative: 1
Best Local Similarity: 89.18% Mismatches: 1
Query Match: 48.03% Indels: 19
Gaps: 1
IS-10-010-050A-2 (1-346) x AL136440 (1-169362)
ly 153 CysPhepHeGluGlyIleaspAspValHsITrplysGluAsnGlyThrLeuValGlnVal 173
lb 118295 TGTGTTTTTAA----- 118306
ly 173 AlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValIysGlnAspAsnGlu 192
lb 118307 -----CTGGAACACATGTTTCAACCAATGGCAAGTGGGTGAACAGACATGAA 118357
ly 193 ThrGlyIleTyrTyrGluThrTrpAsnValIysAlaSerProGluIuysGlyValaGluThr 212
lb 118358 ACAGGAATTTATATGACACATGAGATGTAAAGCCAGCCAGAAAGGGGGCAGAGACA 118417
ly 213 TrpPheAspSerTyrTrpPheCysSerIysPheValIleuArgThrPheAsnIysLeuIaGlu 232
lb 118418 TGGTTTGATTCCTCAACACATGTCCTCAAAATTTGTTTAAAGACCTTAAAGTTGGCTGAA 118477
ly 233 PheGlyAlaGluPheIysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGly 252
lb 118478 TTTGGAGCGAGATTCAAGAACATAGAACCACTATACAGAAATATTTCTTTACAGTGA 118537
ly 253 GluProThrTyrLeuGluIysAsnGluThrSerValPheGlyProThrGlyAsnIysThrLeu 272
lb 118538 GAACCTACTTATCTGGGAATATGAAACATCTGTTTGGGCCCAAGGAACAAGACTCTT 118597

```

```

Qy 273 GlyLeuAlaIleIysArgPheTyrTyrProPheIysProHsIleuProThrIysGluPhe 292
Db 118598 GGTTCACCAATAAABATTTTATTACCCCTTCAACCACTTTGGCACTAAGAAATT 118657
Qy 293 LeuIleuSerLeuIeuGlnIlePheAspAlaValIleValHsIysGlnIleTyrLeuPhe 312
Db 118658 CTGTGAGTCTCTTGCAAATTTTGTGAGTGTGTGCAAAAGCTCTATTTGTTT 118717
Qy 313 TyrAsnDheGluTyrTrpPheLeuProMetIysPheProPheIleIysIleThrTyrGlu 332
Db 118718 TATAATTTGAAATTTGTTTATCTATGAAATTCCTTTATTAATAACATATGAA 118777
Qy 333 GluIleProLeuProIleArgAsnIysThrIleuSerGlyIleu 346
Db 118778 GAAATCCCTTACTATCAAGAACAAACAACTCTGTGTTA 118819

RESULT 14
AL359875/c 251187 bp DNA linear HTG 13-JUN-2001
LOCUS Homo sapiens chromosome 13 clone RP11-453N2, 53 unordered pieces.
DEFINITION AL359875.3 GI:9864226
ACCESSION AL359875.3
VERSION HTG; HTGS PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1
AUTHORS Burton,J.
TITLE Direct Submission
JOURNALS Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 22, 2000 this sequence version replaced gi:8894433.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA453N22
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 220597 bases at least Q40
Consensus quality: 234299 bases at least Q30
Consensus quality: 241103 bases at least Q20
Insert size: 245987; sum-of-contigs
Insert size: 190986; agarose-fp
Quality coverage: 2.52x in Q20 bases; sum-of-contigs Quality
coverage: 3.92x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2130: contig of 2130 bp in length
* 2230: gap of 100 bp
* 2231 5689: contig of 3459 bp in length
* 5690 5789: gap of 100 bp
* 5790 10046: contig of 4257 bp in length
* 10047 10146: gap of 100 bp
* 10147 12624: contig of 2478 bp in length
* 12625 12724: gap of 100 bp
* 12725 20393: contig of 8269 bp in length
* 20394 21093: gap of 100 bp
* 21094 28936: contig of 7843 bp in length
* 28937 29036: gap of 100 bp

```

*	197683	197788:	gap of 100 bp	in length
*	197783	201788:	contig of 4006 bp	in length
*	201789	201888:	gap of 100 bp	in length
*	201889	204060:	contig of 2172 bp	in length
*	204061	204160:	gap of 100 bp	in length
*	204161	206237:	contig of 2077 bp	in length
*	206238	206373:	gap of 100 bp	in length
*	206378	210376:	contig of 4039 bp	in length
*	210377	214765:	gap of 100 bp	in length
*	210377	225245:	contig of 14769 bp	in length
*	225246	225345:	gap of 100 bp	in length
*	225346	230027:	contig of 4682 bp	in length
*	230028	230127:	gap of 100 bp	in length
*	230128	232610:	contig of 2483 bp	in length
*	232611	232710:	gap of 100 bp	in length
*	232711	234942:	contig of 2232 bp	in length
*	234943	240542:	gap of 100 bp	in length
*	235043	240958:	contig of 5516 bp	in length
*	240959	241058:	gap of 100 bp	in length
*	241059	251187:	contig of 10129 bp	in length
FEATURES	Location/Qualifiers			
source	1..251187			
	/organism="Homo sapiens"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:9606"			
	/chromosome="13"			
	/clone="RP11-453N2"			
	/clone="RP11-RP01-11.2"			
misc_feature	1..2130	/note="assembly_fragment:00469		
	fragment_chain:1			
misc_feature	2231..5689	/note="assembly_fragment:01219		
	fragment_chain:1			
misc_feature	5790..10046	/note="assembly_fragment:00352		
	fragment_chain:1			
misc_feature	10147..12624	/note="assembly_fragment:00547		
	fragment_chain:2			
misc_feature	12725..20993	/note="assembly_fragment:00617		
	fragment_chain:2			
misc_feature	21094..28936	/note="assembly_fragment:01248		
	fragment_chain:2			
misc_feature	29037..32230	/note="assembly_fragment:00606		
	fragment_chain:3			
misc_feature	32331..35153	/note="assembly_fragment:01016		
	fragment_chain:3			
misc_feature	35254..37649	/note="assembly_fragment:01277		
	fragment_chain:3			
misc_feature	37750..40245	/note="assembly_fragment:00920		
	fragment_chain:4			
misc_feature	40346..48470	/note="assembly_fragment:01288		
	fragment_chain:4			
misc_feature	48521..52169	/note="assembly_fragment:00615		
	fragment_chain:4			
misc_feature	52270..56443	/note="assembly_fragment:00009		
	fragment_chain:5			
misc_feature	56544..63088	/note="assembly_fragment:00448		
	fragment_chain:5			
misc_feature	63189..65281	/note="assembly_fragment:00027		

Segment Scores:	4.24e-72	Length:	251187
id. No.:	925.50	Matches:	173
Percent Similarity:	89.69%	Conservative:	1
Local Similarity:	89.18%	Mismatches:	1
Match:	48.03%	Indels:	19
	2	Gaps:	1
-10-010-050a-2 (1-346) x AL359875 (1-251187)			
153 CysPhePheGluGluGlyLeuAspAspValHisTyrPheGluAsnGlyThrLeuValGlnVal	172		
194107 TGGTTTTTTTAA-----	194696		
173 AlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValIlysglnAspAsnGlu	192		
194095 -----CTAGGAAACATGTTCAACCAATCGCAAACTGGGTGAAACAGCAATGAA	194645		
193 ThrGlyIleTyrTyrGluThrTyrPheValIleValIleSerProGluIlyPheGluIleGluThr	212		
194044 ACAGGAAATTTATATGACATGAAATGTAAGCCAGCCCGAAGAGGGGCGAGAGACA	193685		
213 TrpPheAspSerTyrAspCysSerIlyPheValIleuArgThrPheAsnIlysglnIleuIle	232		
193984 TGGTTGATTCCTACGACCTGTTCCAAATTTGTTTAAAGACCTTTAAACAAGTTGGCTGAA	193625		
233 PheGlyAlaGluPheIlyAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGly	252		
193924 TTTGGAGCAGAGTTCAAGAACATGAAACCACTAACAAGATATTTCTTTACAGTGA	193865		
253 GluProThrTyrIleuGluAsnGluThrSerValPheGlyProThrGlyAsnIlysglnIleu	272		
193864 GAACCTCACTTATCGGAAATGAAACATCTGTTTGGGCCAAGAGAAACAGACTCTT	193805		
273 GlyLeuAlaIleIlyAsnGlyPheTyrTyrProPheIlyProHisLeuProThrIlysglnIleu	292		
193804 GGTTCGCCATTAAGATATTTTACCCCTTAAACCAATTTGCCAATTAAGATATTT	193745		
293 LeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisIlysglnIlePheTyrLeuPhe	312		
193744 CTGTGAGTCTCTGCAAAATTTTGATGACAGTATTTGCAAAACAGTTCTATTTGTTT	193685		
313 TyrAsnPheGluTyrTyrPheLeuPrometIlyPheProPheIleIlyIleThrTyrGlu	332		
193684 TATPATTTTAAATTTGTTTACCATGAAATTCCTTTATTAATTAACATATGAA	193625		
333 GluIleProLeuProIleArgAsnIlyThrLeuSerGlyLeu	346		
193624 GAATTCCTTACCTATCAGAAACAAACACCTCTGCTTAA	193583		
SPLIT 15			
014175_0			
COMMENT			
Sequence split into 4 fragments	LOCUS AE014175 Accession AE014175		
Fragment Name	Begin	End	
AE014175_0	1	110000	
AE014175_1	100001	210000	
AE014175_2	200001	310000	
AE014175_3	300001	404829	
CUS	AE014175	404829 bp	DNA linear ROD 06-AUG-2002
FINITION	Mus musculus plebeald deletion region section 3 of 11 of the		
complete sequence.			
CESSION	AE014175 AE013600		
RSION	AE014175.1 GI:22128029		
YMORDS			
ORCE	Mus musculus (house mouse)		
ORGANISM			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurgnath; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 404829)		
AUTHORS	Peterson, K., King, B., Hagge-Greenberg, A., Roix, J., Bult, C. and		
	O'Brien, T.		
TITLE	Functional and comparative genomic analysis of the plebeald deletion		

JOURNAL	region of mouse chromosome 14
MEDLINE	Genomics 80 (2), 172 (2002)
PUBMED	22150871
REFERENCE	12160731
AUTHORS	2 (bases 1 to 404829)
TITLE	Adams, M. and Mural, R.
JOURNAL	Direct Submission
	Submitted (31-MAY-2002) Celera Genomics, 45 West Gude Drive,
	Rockville, MD 20850, USA
	The strings of n's in this record represent gaps between contigs,
	and the length of each string corresponds to the length of the gap
	except where there are 20 n's. A string of 20 n's represents a gap
	of unknown length.
FEATURES	
source	Location/Qualifiers
	1..404829
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="14"
	/map="Mt164-Mt165"
	<38137..>47216
	/gene="Irg1"
	<38137..>47216
	/gene="Irg1"
	<38137..>47216
	/gene="Irg1"
	/product="immune-responsive gene 1"
	/note="partial transcript sequence derived from alignment
	to GenBank Accession Number L38281"
	<61259..>68607
	/gene="Cln5"
	join((61259..61421,62721..62886,64274..64499,
	66916..>68607)
	/gene="Cln5"
	/product="ceroid-lipofuscinosis, neuronal 5"
	/note="partial transcript derived from alignment to
	tentative consensus sequence TC227789, from TIGR Mouse
	Gene Index (v. 6.0)."
	complement(<73858..>86476)
	/gene="Fbl3a"
	complement(join(<73858..74501,80293..80464,83295..83417,
	86126..>86476))
	/gene="Fbl3a"
	/product="leucine-rich repeat-containing F-box protein
	Fbl3a"
	complement(join(73858..74501,80293..80464,83295..83417,
	86126..>86476))
	/gene="Fbl3a"
	/codon_start=1
	/product="leucine-rich repeat-containing F-box protein
	Fbl3a"
	/protein_id="AAW92567.1"
	/db_xref="GI:22128030"
	/translation="RMKRGGRSDSDSAEEGTAEKRPRTTOERSQPCDMKNTLQDI
	VHVEFYPLDRAHSQYCRNNOVFHNPDIIMRCEPFLNPATSYIAKATPELIKQ
	IKRSHNLQYVSFKVDSKESAEACDLISQVNSLTKLGLISTARSFMDLPESH
	FISALIVVFNVSKEISLSKIDTPVDDPDLKIVANNSTDLKLKNSCPHVSPAIL
	CVADQHGRLRELVHLLSDELLALASEKVRLEHLIDVSENPQGHPTTQKS
	SWDAIKHSKPKNIWVFLVEEEDPPRPRIPTHLYEGRSVSDVGRVMTQCPR
	LVELVYCGALRPDLBELRIARCKNLSAIGLGECEVCSAFVEYVNMCGRLSGLS
	IMEEVLIPDKTSLLEQIMHVESKHGKRVFPMDFW"
	complement(<106136..>339446)
	/gene="Prl1"
	complement(<106136..>339446)
	/gene="Prl1"
	/product="Pam/hw/rpm-1"
	/note="PAM, protein associated with Myc; partial
	transcript derived from alignment of Homo sapiens NCBI
	RefSeq sequence NM_015057"
ORIGIN	
Alignment Scores:	
Pred. No.:	1..11e-54
Score:	730.00
Percent Similarity:	87.65%
	Length: 110000
	Matches: 127
	Conservative: 22

Best Local Similarity: 74.71% Mismatches: 21  
Query Match: 37.88% Indels: 0  
DB: 10 Gaps: 0

IS-10-010-050a-2 (1-346) x AE014175\_0 (1-110000)

```

  177  GlysMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGlnThrGlyIleTyr 196
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
  66915  GGAACACATTTTAAACAAGTGGCCGAGTGGTGAAGCAGACATGAATCGGATTTAT 66974
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  197  TyrGlnThrTrpAsnValLysAlaSerProGluLysGlyAlaGlnThrTrpPheAspSer 216
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  66975  TATGAGACATGAGCGGCGCGAGCCGACAGACAGGGGCCAGACGCGTTCGAGTCC 67034
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  217  TyrAspCysSerLysPheValLeuArgThrPheAsnLysLeuAlaGlnPheGlyAlaGln 236
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  67035  TACGACTGTTGCAATTTGCTTAAAGACATATAGAAATGGCTGAATTTGAAACAGA 67094
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  237  PheLysAsnIleGlnThrAsnTyrThrArgLepheLeuTyrSerGlyLysProThrTyr 256
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  67095  TTCAAGACATGAAACAAACTATACGAAATATTTCTTACAGTGAAGCCTATTAC 67154
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  257  LeuGlyAsnGlnThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle 275
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  67155  CTGGAAATGAAACATCTATTGCGGCCCAAGGAAACAGACTCTGCTTGGCCATA 67214
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  277  LysArgPheTyrTyrProPheLysProHisLeuProThrLysGlnPheLeuLeuSerLeu 295
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  67215  AAAAATTTATGCCCCCTTCAGACCGATTTGTCAACCAAGATTTCTGATGAATTC 67274
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  297  LeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGln 315
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  67275  TTGAAATTTTGTATCAGTATTATACACAGACAGTTCCTACTGTTTATACCTTGAG 67334
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  317  TyrTrpPheLeuProPheLysPheProPheIleLysIleThrTyrGlnGluIleProLeu 336
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  67335  TATTGGTTCTACCAATGAACCCCTTGTCTCAAAATACATACAGAAACCCGCTTA 67394
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  337  ProIleArgAsnLysThrLeuSerGlyLeu 346
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  67395  CCTACCGACATACACATTTACCGACTTG 67424
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

Search completed: April 25, 2004, 03:18:12  
Job time : 6841.36 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame\_plus\_p2n model

un on: April 24, 2004, 22:59:32 ; Search time 5973.28 Seconds

(without alignments)  
2292.945 Million cell updates/sec

file: US-10-010-050A-2\_COPY\_31\_346  
erfect score: 1758  
equence: 1 RVSGTSPRRKMPVYKRFDF.....IKTYEIRIPINKTUSGL 316

coring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

sarched: 3470272 seqs, 21671516995 residues

ctal number of hits satisfying chosen parameters: 6940544

imum DB seq length: 0  
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

ommand line parameters:

MODEL=frame+ p2n model -DEV=x1h  
2=/cgn2.1/USPRO.spool/US10010050/runat.22042004.113204.27549/app.query.fasta\_1.1372  
DB=genembl -GEMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0  
UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
OUTMT=ptco -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
USER=US10010050 @CGN 1 1 7433 @runat.22042004.113204.27549 -NCPU=6 -ICPU=3  
NO\_MMAP -LARGECQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_pi.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_ats.\*  
12: gb\_by.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_ov.\*  
22: em\_or.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rod.\*  
36: em\_htg\_man.\*  
37: em\_htg\_vtr.\*  
38: em\_by.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1758	100.0	1486	BD073402	BD073402 Secreted
2	1758	100.0	4080	AF068227	AF068227 Homo sapi
3	1747	99.4	1751	BD136338	BD136338 95 human
4	1744	99.2	2120	BD127858	BD127858 Primer fo
5	1744	99.2	2120	AK075109	AK075109 Homo sapi
6	1515	86.2	1038	BD073413	BD073413 Secreted
7	1423	80.9	2318	BC025487	BC025487 Mus muscu
8	1396	79.4	198524	AC109554	AC109554 Rattus no
9	1339	76.2	227920	AC107510	AC107510 Rattus no
10	925.5	52.6	106988	AC001226	AC001226 Genomic s
11	925.5	52.6	169362	AL136440	AL136440 Homo sapi
12	925.5	51.6	221187	AL1359875	AL1359875 Homo sapi
13	896	51.0	697	BD125187	BD125187 Primer fo
14	896	51.0	697	BD126453	BD126453 Primer fo
15	730	41.5	110000	AE014175_0	AE014175 Mus muscu
16	730	41.5	156823	AC102815	AC102815 Mus muscu
17	727	40.4	506	BD060465	BD060465 Secreted
18	716	40.7	229015	AC106966	AC106966 Rattus no
19	711	40.4	217336	AC131344	AC131344 Rattus no
20	617.5	35.1	251187	AL359875	AL359875 Homo sapi
21	582	33.1	198524	AC109554	AC109554 Rattus no
22	272.5	15.5	258273	AC113855	AC113855 Rattus no
23	266	14.6	110000	AC116305_0	AC116305 Dictyoste
24	241	13.7	473	AX333881	AX333881 Sequence
25	241	13.7	473	AX335993	AX335993 Sequence
26	146.5	8.3	256879	AC116982	AC116982 Dictyoste
27	111.5	6.3	223182	AL929049	AL929049 Zebrafish
28	104.5	5.9	2162	BC014298	BC014298 Homo sapi
29	104.5	5.9	2175	AX881019	AX881019 Sequence
30	104.5	5.9	2175	BD158729	BD158729 Primer fo
31	104.5	5.9	2175	AK023964	AK023964 Homo sapi
32	103	5.9	1634	BD275044	BD275044 50 Human
33	103	5.9	2904	AK075480	AK075480 Homo sapi
34	103	5.9	171109	EX005355	EX005355 Zebrafish
35	101	5.7	1638	AY198311	AY198311 Streptoco
36	101	5.7	1946	AF407190	AF407190 Trichophy
37	100	5.7	1973	AF407186	AF407186 Trichophy
38	100	5.7	55686	AC095291_3	AC095291_3 of
39	100	5.7	110000	AC095291_2	AC095291_2 of
40	100	5.7	165504	AC140352	AC140352 Mus muscu
41	100	5.7	175765	AL683871	AL683871 Human DNA
42	100	5.7	219318	AC133392	AC133392 Rattus no
43	100	5.7	251506	AC106702	AC106702 Rattus no
44	99.5	5.7	305961	AE016937	AE016937 Bacteroid
45	99	5.6	140332	AC118538	AC118538 Felis cat

RESULT 1

## ALIGNMENTS



BD073402 1486 bp DNA linear PAT 27-AUG-2002  
 LOCUS Secreted protein which human chromosome 13 encodes.  
 DEFINITION  
 ACCESSION BD073402 GI:22619005  
 VERSION BD073402.1  
 KEYWORDS JP 2001511345-A/1.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 1486)  
 AUTHORS Sheppard, P.O. and Gilbertson, D.G.  
 TITLE Secreted protein which human chromosome 13 encodes  
 JOURNAL Patent: JP 2001511345-A 1 14-AUG-2001;  
 ZYMOGENETICS INC  
 COMMENT OS Homo sapiens (human)  
 PN JP 2001511345-A/1  
 PD 14-AUG-2001  
 PE 24-JUL-1998 JP 200504249  
 PR 24-JUL-1997 US 60/053613  
 PI PATL O SHEPPARD, DIBRA G GILBERTSON  
 PC C12N15/09; A61K38/00; A61K48/00; C07K14/47; C07K16/18; C12N1/15; PC C12N1/19;  
 PC C12N1/21; C12N5/10; C12P21/02; C12Q1/68; C12N15/00; A61K37/02; C12N5/ PC 00  
 CC Secreted protein which human chromosome 13 encodes FH Key  
 FT CDS Location/Qualifiers  
 1..1486  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,95e-170 Length: 1486  
 Score: 1758.00 Matches: 316  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0  
 JB: 6  
 US-10-010-050a-2\_COPY\_31\_346 (1-316) x BD073402 (1-1486)  
 QY 1 ArgValSerGlyLeuProSerArgHisTyrProValProTyrTyrLysArgPheAspPhe 20  
 DB 137 CGGGTCTCGGGGCAATCCCTCCCGCGCCACTGGCCGGTGCCTTACAAAGCCCTTGAAGCTTC 196  
 QY 21 ArgProLysProAspProTyrTyrGlnAlaLysTyrPheCysProThrGlySerPro 40  
 DB 197 CGTCAAAACCTGATCTTATGTCAGAGCTATGATCTTCTGCTCCACTGCTACCT 256  
 QY 41 LLeuProValMetGluGlyAspAspAspLysLysLysLysLysLysLysLysLysLys 60  
 DB 257 ATCCCACTTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 316  
 QY 61 GluPheLysTyrGlyAspLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80  
 DB 317 GAATTTAAATATGAGACCTCCCTGGGACACTTGAATTTATGCAATGATGATGATGATGATG 376  
 QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGly 100  
 DB 377 AGAAGTACATTAATCGGCAAGACATCAACATGAAATGATGAAATGAAATGAAATGAAATG 436  
 QY 101 AsnGlyThrPheProHisLysLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120  
 DB 437 AACTGTACATTTCCCATCTCCGACCTGAATGATGATGATGATGATGATGATGATGATGATG 456  
 QY 121 AAlaLysCysPhePheGlnGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 140  
 DB 497 GCTGCTGCTTTTGTGAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 556

QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrProValLysGlnAsp 160  
 DB 557 CAAGTAGCAATATATATGAGAAACATGTTCAACCAATAGCCAAAGTGGGTAAACAGAC 616  
 QY 161 AsnGluThrGlyLysTyrTyrGluThrTyrAsnValLysLysLysLysLysLysLysLys 180  
 DB 617 AATGAAACAGAAATTTATATGAGACATGAAATGAAATGAAATGAAATGAAATGAAATG 676  
 QY 181 GluThrTyrPheAspSerTyrAspCysSerLysPheValLysArgThrPheAsnLysLeu 200  
 DB 677 GAGACATGCTTTGATCTCTTACGACTGTTCCAAATTTGTTGTTAAGACCTTTTACAGTTG 736  
 QY 201 AlaGluPheGlyAlaGluPheLysAsnLysLysLysLysLysLysLysLysLysLysLys 220  
 DB 737 GCTGATTTTGGAGAGAGATTCAGAAACATGAAACCACTATACAGATATATTTCTTTAC 796  
 QY 221 SerGlyLysProThrTyrLysGlyAsnGluThrSerValPheGlyProThrGlyLysLys 240  
 DB 797 AGTGAGAACCTACTTACTGGAATGAAACATCTGTTTGGGCGCAACAGAAACAG 856  
 QY 241 ThrLeuGlyLeuAlaLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260  
 DB 857 ACTCTGCTTTAGCCATTAATAAGATTTTATACCCCTTCAACACCATTTGCCAAGCTAA 916  
 QY 261 GluPheLeuLeuSerLeuLeuGlnLysPheAspAlaValLysLysLysLysLysLysLys 280  
 DB 917 GAATTTCTGTTGAGCTCTCTGCAAAATTTTATGATGATGATGATGATGATGATGATGAT 976  
 QY 281 LeuPheTyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheLysLysLys 300  
 DB 977 TTGTTTAAATTTGAAATATGTTTATCTTACCTATGAAATTTCCCTTTATTTAAATTA 1036  
 QY 301 TyrGluGluLysLeuProLeuProLysArgAsnLysThrLysSerGlyLeu 316  
 DB 1037 TATGAAAGAAATCCCTTACCTATGAAACAAACAAACAACTCTGCTTTA 1084  
 RESULT 2  
 AF068227 4080 bp mRNA linear PRI 28-JUL-1998  
 LOCUS  
 DEFINITION Homo sapiens putative transmembrane protein (CLN5) mRNA, complete cds.  
 ACCESSION AF068227  
 VERSION AF068227.1 GI:3342385  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 4080)  
 AUTHORS Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.  
 TITLE CLN5, a novel gene encoding a putative transmembrane protein mutated in Finnish variant late infantile neuronal ceroid lipofuscinosis  
 JOURNAL Nat. Genet. 19 (3), 286-288 (1998)  
 MEDLINE 98324783  
 PUBMED 9662406  
 REFERENCE 2 (bases 1 to 4080)  
 AUTHORS Savukoski, M., Klockars, T., Holmberg, V., Santavuori, P., Lander, E.S. and Peltonen, L.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-MAY-1998) Department of Human Molecular Genetics, National Public Health Institute, Mannerheimintie 166, Helsinki 00300, Finland  
 FEATURES  
 source 1..4080  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /map="13q22"  
 gene 1..4080

CDS

```

/gene="C1N5"
1293..2515
/gene="C1N5"
/codon_start=1
/product="putative transmembrane protein"
/protein_id="AAC27614.1"
/db_xref="ncbi:3342386"
/translation="MRRNLKGPSSGADNQCQCAPRPGCLAPRHLPPASQASGSGS
TGCSTMAQEDVDAQEMRKGARAGASCMALALMTLVVPGMSKVSGLPSRRM
PVVYRPRPRKPDYCOAKYIPCTGSPIPVMEGDDIDIEVRLQAPVBEKYDILG
HLKIHDAIGRSLTGKNTYMEVLEFQLNCTEPHLRPMQAPFMCQAAQFEG
IDDHMKENGTLVQVATISGMENOMAKVQNDNETGIYETVAVKASPEKAGTFWD
SYDSKFLTRFNKLAEFGARFKNTETWTETIFLYSGSPYTGNTVKSFGSGTKTG
LAIKRFYPERKHLPTKEFLSLDQIFDAVIVHQFYLFTFEPVFLMKRPFILKITY
EELPLIRNKTLGL"

```

IGIN

Alignment Scores:

```

ed. No.: 1,09e-169 Length: 4080
Score: 1758.00 Matches: 316
Percent Similarity: 100.00% Conservative: 0
Percent Local Similarity: 100.00% Mismatches: 0
Very Match: 100.00% Indels: 0
Gaps: 0

```

1-10-010-050a-2\_copy\_31\_346 (1-316) x AF068227 (1-4080)

```

1 ArgValSerGlyIleProSerArgArgHisTyrProValProTyrIleArgPheAspPhe 20
1566 CCGGCTGGGGGATCCCTCCGGCGGCACTGGCGGTGCTTCAACAGCGTTTACCTTC 1625
21 ArgProIleProAspProTyrCysGlnAlaIleTyrIlePheCysProThrGlySerPro 40
1626 CGTCCAAAACCGATCCCTTATGTCACGTAGTATCTTCTGTCACACGCGCCACCT 1685
41 IleProValMetGlnGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60
1686 ATCCCAAGTTATGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1745
61 GluPheIleTyrGlyAspIleuLeuGlnIleIleuIleMetHisAspAlaIleGlyPhe 80
1746 GAATTTAAATATGAGACCTCTCGGAGACCTGAAATATATCATGATGATGATGATGATG 1805
81 ArgSerThrLeuThrGlyIleValAsnTyrThrMetGluTyrIleGluLeuPheGlnLeuGly 100
1806 AGAAGTACATTAACATCGGACAGAACACCAAGATGATGATGATGATGATGATGATGATG 1865
101 AsnCysThrPheProHisIleuAsnArgProGluMetAspAlaProPheThrCysAsnGlnGly 120
1866 AACTGTACATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGATGATG 1925
121 AlaAlaCysPhePheGlnGlyIleAspAspValIleTyrIleGluAsnGlyThrLeuVal 140
1926 GCTGCTGCTGTTTTTGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1985
141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValIleGlnAsp 160
1986 CAAGTAGCAACTATATACAGAAACATGTTCCAAATGGCAAAAGTGGTGAACAGAGAC 2045
161 AsnGluThrGlyIleTyrIleGluThrPheAsnValIleValSerProGluIleGlyAla 180
2046 AATGAAACAGGAATTTATATATAGACATGAAATGTAAGAACCAAGCCCAAGAAAGGGGCA 2105
181 GluThrTyrPheAspSerTyrAspCysSerIlePheValIleuArgThrPheAsnIleu 200
2106 GAGACATGTTTGAATCTTACACACTGTTCCAAATTTGTTAAGGACCTTTTAAACAAGTTG 2165
201 AlaGluPheGlyAlaGluPheIleAsnIleGluThrThrAsnTyrThrArgIlePheLeuTyr 220
2166 GGTGAATTTGGAGCGAGAGTTCAAGAACATTAAGAACCAATTAACAAGAAATATTTCTTAC 2225
221 SerGlyGluProThrTyrIleuGlyAsnGluThrThrSerValPheGlyProThrGlyAsnIle 240

```

```

Db 2226 AGTGAGAACCTACTATATCTGGAATGAAACATCTGTTTTTGGGCCACAGAAACAG 2285
Oy 241 ThrLeuGlyLeuAlaIleIleValArgPheTyrTyrProPheIleProHisIleuProThrIle 260
Db 2286 ACTCTTGTTTGAAGCAAAAAGATTTATTAATCCCTTCAACACACATTTGCCAATGAA 2345
Oy 261 GluPheIleuSerLeuLeuGlnIlePheAspAlaValIleValIleIleGlnIlePheTyr 280
Db 2346 GAATTTCTGTTGAGTCTCTTGCAAAATTTTTCATGACAGATTTGACCAACACAGTTCTAT 2405
Oy 281 LeuPheTyrAsnPheGluTyrTyrPheLeuProMetIlePheProPheIleIleThr 300
Db 2406 TTGTTTATATATTTTGAATATGATGTTTATACCTATGAATATCCCTTTATTAATAATA 2465
Oy 301 TyrGluGluIleProLeuProIleArgAsnIleThrIleuSerGlyLeu 316
Db 2466 TATGAGAAATCCCTTATCCTATCAGAAACAAACACACTCTGTTT 2513

```

RESULT 3

BD136338

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1751)
Ruben, S.M., Ni, J., Rosen, C.A., Yu, G.L., Young, P.E., Fen, P.,
Soppet, D.R., Wei, Y.F., Endress, G.A., Dunn, R.D., Kyaw, H., Ebner, R.,
Lafleur, D.W., Olsen, H.S., Shi, Y. and Moore, P.A.
95 human secretory proteins
Patent: JP 2002506627-A 25 05-MAR-2002;
HUMAN GENOME SCIENCES INC
OS Homo sapiens (human)
PN JP 2002506627-A/25
PD 05-MAR-2002
PF 18-MAR-1999 JP 2000536733
PR 19-MAR-1998 US 60/078566, 19-MAR-1998 US 60/078576 PR
19-MAR-1998 US 60/078573, 19-MAR-1998 US 60/078574 PR
19-MAR-1998 US 60/078579, 19-MAR-1998 US 60/078578 PR
19-MAR-1998 US 60/078581, 19-MAR-1998 US 60/078577 PR
19-MAR-1998 US 60/078563, 01-APR-1998 US 60/080314 PR
01-APR-1998 US 60/080312, 01-APR-1998 US 60/080313 PI
STEVEN
M RUBEN, JIAN NI, CRAIG A ROSEN, GDO
LIANG YU, PAUL B YOUNG,
PING FENG,
DANIEL R SOPPET, YING FEI WEI, GREGORY A ENDRESS, ROXANNE D DUAN,
HUA KIAN,
REINHARD EBNER, DAVID W LAFLUR, HENRIK S OLSEN, YANGGU SHI, PAUL
A MOORE
PC C12N15/09, A61K38/00, A61K48/00, A61P43/00, C07K14/47, C07K16/18,
PC C12N1/15,
PC C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12Q1/68, G01N33/53, G01N33/
566,
PC C12N15/00, A61K37/02, C12N5/00
CC n equals a, t, g, or c
CC n equals a, t, g, or c
CC n equals a, t, g, or c
CC n equals a, t, g, or c
CC n equals a, t, g, or c
CC n equals a, t, g, or c
CC n equals a, t, g, or c
FH key
FT source
1. 1751
/organism="Homo sapiens (human)"
/feature
1. 1751
/location/Qualifiers
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

FEATURES

source

## ORIGIN

## Alignment Scores:

Pred. No.: 4,88e-169 Length: 1751  
 Score: 1747.00 Matches: 314  
 Percent Similarity: 99.37% Conservative: 2  
 Best Local Similarity: 99.37% Mismatches: 0  
 Query Match: 99.37% Indels: 0  
 DB: 6 Gaps: 0

US-10-010-050A-2\_COPY\_31\_346 (1-316) x BD136338 (1-1751)

2y 1 ArgValSerGlyIleProSerArgHISrPrProValProTyrLysArgPheAspPhe 20  
 Db 139 CGGGTCTCGGGCAATCCCTCCGGCCGCTGCGCGCTTGAAGGCTTGAATTC 193  
 2y 21 ArgProLysProAspProTyrCyGlnAlaLysTyrThrPheCysProThrGlySerPro 40  
 Db 199 CGTCCAAAACCTGATCTTATGTCTCAAGCTAAGTAACTTCTGTCCAACTGGCTCACT 258  
 2y 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60  
 Db 259 ATCCCACTTATGAGGGGTGATGATGACATTGAAGTTTTCATTACAGCCCAAGTATGG 318  
 2y 61 GluPheLysTyrGlyAspLeuLeuGlyHISLeuLysIleMetHISAspAlaIleGlyPhe 80  
 Db 319 GAATTTAAATATGAGAGCTCTCGGACACTTGAATTAATGATGATGCTTGAATTC 378  
 2y 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrGlyLeuPheGlnLeuGly 100  
 Db 379 AGAGTACATTAACTGCAAGAACTACACAAATGATGATGATGATGATGATGATGATGAT 438  
 2y 101 AsnCySerThrPheProHISLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120  
 Db 439 AACGTACATTTCCTCCATCTCCGACCTGAAATGATGATGATGATGATGATGATGATGAT 458  
 2y 121 AlaAlaCysPhePheGlnGlyIleAspAspValHISrPrLysGlnLeuVal 140  
 Db 499 GCTGCTGCTCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 558  
 2y 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrProValLysGlnAsp 160  
 Db 559 CAAGTAGCACTAATATCAGAAACATGTTCAACCAATGGGAAATGGGAAACAGAGAC 618  
 2y 161 AsnGluThrGlyIleTyrTyrGluThrTyrAsnValLysAsnSerProGluLysGlyAla 180  
 Db 619 AATGAAACAGAAATTTATATGACATGATGATGATGATGATGATGATGATGATGATGAT 678  
 2y 181 GluThrTyrPheAspSerTyrAspCysSerLysPheValLysArgThrPheAsnLysLeu 200  
 Db 679 GAGCATGCTTGAATCTCTACGACTGTTCCAAATTTGTGTTAAGGACCTTTAAACAAGT 738  
 2y 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220  
 Db 729 GGTGAATTTGAGAGCAGAGCTTCAGAAACATAGAACCACTATACARGAATATTTCTTAC 798  
 2y 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240  
 Db 799 AGTGAGAGAACCTAATCTGGAATGAAATGAAATGATGATGATGATGATGATGATGATGAT 858  
 2y 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeuProThrLys 260  
 Db 859 ACTCTGCTTATGACATATAAAGATTTTATATACCTTCAAAACCAATTTGGCACTTAA 918  
 2y 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHISLysGlnPheTyr 280  
 Db 919 GAATTTCTGTTGAGCTCTCTGCAAAATTTTATGATGAGATGATGATGATGATGATGATGAT 978  
 2y 281 LeuPheTyrAspPheGluTyrTyrPheLeuProMetLysPheProPheIleLysIleThr 300  
 Db 979 TTGTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1038  
 2y 301 TyrGluGluIleProLeuProIleArgAsnLysrPrLeuSerGlyLeu 316

Db 1039 TATGAGAAATCCCTTATCTATCTAGAAACAAACACTCTCTGTTTAA 1086

RESULT 4  
 LOCUS BD127858 2120 bp DNA linear PAT 18-SEP-2002  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION BD127858  
 VERSION BD127858.1 GI:23222803  
 KEYWORDS JP 2002017375-A/3289.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 2120)  
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,  
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and  
 Koga,H.

TITLE Primer for synthesizing full-length cDNA and use thereof  
 JOURNAL Patent: JP 2002017375-A 3289 22-JAN-2002;  
 HELIX RESEARCH INSTITUTE

## COMMENT

OS Homo sapiens (human)  
 PN JP 2002017375-A/3289  
 PD 22-JAN-2002 JP 2002053172  
 PF 07-JUN-2000 JP 2002053172  
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO  
 PI ISHII,  
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI  
 SHINICHI KOJIMA,  
 PI TETSUO OTSUKI, HISASHI KOGA

PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,  
 C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 Location/Qualifiers  
 (16)..(1089).  
 FT CDS

FEATURES  
 source 1..2120  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.27e-168 Length: 2120  
 Score: 1744.00 Matches: 314  
 Percent Similarity: 99.37% Conservative: 0  
 Best Local Similarity: 99.37% Mismatches: 2  
 Query Match: 99.20% Indels: 0  
 DB: 6 Gaps: 0

US-10-010-050A-2\_COPY\_31\_346 (1-316) x BD127858 (1-2120)

2y 1 ArgValSerGlyIleProSerArgHISrPrProValProTyrLysArgPheAspPhe 20  
 Db 142 CGGGTCTCGGGCAATCCCTCCGGCCGCTGCGCGCTTGAAGGCTTGAATTC 201  
 2y 21 ArgProLysProAspProTyrCyGlnAlaLysTyrThrPheCysProThrGlySerPro 40  
 Db 202 CGTCCAAAACCTGATCTTATGTCTCAAGCTAAGTAACTTCTGTCCAACTGGCTCACT 261  
 2y 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60  
 Db 262 ATCCCACTTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 321  
 2y 61 GluPheLysTyrGlyAspLeuLeuGlyHISLeuLysIleMetHISAspAlaIleGlyPhe 80  
 Db 322 GAATTTAAATATGAGAGCTCTCGGACACTTGAATTAATGATGATGATGATGATGATGAT 381  
 2y 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrGlyLeuPheGlnLeuGly 100  
 Db 382 AGAGTACATTAACTGCAAGAACTACACAAATGATGATGATGATGATGATGATGATGAT 441

101 AsnCythrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120  
442 AACTGTACATTTCCCATCTCCGACCTGAATGGATGCCCTTCTGGTGTATCAAGC 501  
121 AlaIaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuVal 140  
502 GCTGCTCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 561  
141 GluValAlaThrIleSerGlyAsnMetPheAsnGluMetAlaLysTrpValLysGluAsp 160  
562 CAAGTAGCAACTATATCAAGAAACATGTTTCAACCAATGGCAAGTGGGGAACAGAGC 621  
161 AsnGluThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180  
622 AATGAAACAGCAATTTATTTAGACATGATGTAAAGCCACCCAGAAAGGGGGCA 681  
181 GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200  
682 GAGACATGGTTGATTTCCACGACGTGTCCAAATTTGGTTAAAGACCTTTAACAAATGG 741  
201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTrpTrpArgGilePheLeuTyr 220  
742 GGTGAATTTGGAGCAGAGTTCAAGAACATGAACCAACTTACAGAAATATTTCTTAC 801  
221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240  
802 AGTGAAGAACATATTATCTGGGAATGAACATCTGTTTGGGCCCAACGAAACAG 861  
241 ThrLeuGlyLeuAlaIleLysArgPheTyrTrpProPheLysProHisLeuProThrLys 260  
862 ACTCTTGTTTGGCCATPAAAAAGATTTATTAACCTTCAACCAACATTTGCCAATCAA 921  
261 GluPheLeuLeuSerLeuGlnIlePheAspAlaValIleValHisLysGluPheTyr 280  
922 GAATTTCTGTGAGTCTCTTGCATTTTGTATGACAGTATGGCACAACAGTTGAT 981  
1 281 LeuPheTyrAsnPheGluTyrTrpPheLeuProLysPhePheProPheIleLysIleThr 300  
982 TTGTTTATATTTGAATATGGTATTTAACCTATGAATATCCCTTTATTTAAATATACA 1041  
301 TyrGluGluIleProLeuProIleArgAsnLysTrpLeuSerGlyLeu 316  
1042 TATGAGAAATCCCTTACCTATCAAGAACAAACACTCTCTGGTTTA 1089

ACCS	AK075109	2120 bp	mRNA	linear	PI 03-SEP-2002
DEFINITION	Homo sapiens cDNA FL390628 fis, clone FLACEL003407, highly similar to Homo sapiens putative transmembrane protein (CLN5) mRNA.				
ACCESSION	AK075109				
VERSION	AK075109.1	GI:22760983			
KEYWORDS	oligo capping; fis (full insert sequence) .				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1				
	Isogai, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T., Sugiyama, T., Suzuki, Y., Nagai, K., Sugano, S., Ishii, S., Kawai-Hio, Y., Saito, K., Yamamoto, T., Makamatsu, A., Nakamura, Y., Kojima, S., Nagahara, K., Masubo, Y., Ono, T., Okano, K., Yoshikawa, Y., Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and Nimomiyu, K.				
TITLE	NEDO human cDNA sequencing project				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2120)				
AUTHORS	Isogai, T. and Otsuki, T.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yama, Kisarazu, Chiba 292-0812, Japan (E-mail:genomict@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3985)				
COMMENT	NEDO human cDNA sequencing project supported by Ministry of				

Economy, Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction; Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure, Human genome Center; cDNA 5' - & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.).

**FEATURES**  
**source**

```

source
1. .2120
/organism="Homo sapiens"
/mol_type="mRNA"
/dn_xref="taxon:9606"
/clone="P1ACB1003407"
/tissue_type="placenta"
/clone_lib="P1ACB1"
/note="Cloning vector: pME18sFL3

```

ORIGIN

**Alignment Scores:**

Pred. No.:	1,276-168	Length:	2160
Score:	1744.00	Matches:	314
Percent Similarity:	99.37%	Conservative:	0
Best Local Similarity:	99.37%	Mismatches:	2
Query Match:	99.20%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050A-2 COPY 31 346 (1-316) X AK075109 (1-2120)

QY 1 ArgValIserGlyIleProSerArgArgHisTrpProValProTyrLysArgPheAspPhe 20

Db 142 CGGCTCTCGGCATCCCCTCCCGCGCCACTGGCCGGTGCCTGCAAGCCTTGA

21 AcPrGlyValProAspProTyrCysGlnAlaIleValTyrThrPheCysProThrglySerPro 40

26

100

41 11EFIOALNELCULU8YX0SPHNP9LCCD0VATTCOM7FESTACNCL

D5 262 ATCCAGTATGAGGGTGAATGACACTGAGTTTTCGATACACGCCCCGATTCG 32

61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 80

Db 322 GAATTTAATA TGAGACCTCTGGACACTTGA AATTATGCATGATGCCATTGGATTC 38

81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 10

183 A C T A C T T A A T G G C A A C T A C A C A T G G A A T G G T A T G A C T T T T C C A C T T G C 44

...-DacG]uWetAnoG]aProDhetTymCveAsnG]nG]v 12

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

[illegible]

121 Alaalacybnehegugyileabpvaibisilpyssubmoel mmooua

Db 502 GCTGCCTGCTTTTGAAGGAATGATGATGTTCACTGGAAGGAATGGACATTAAT 56

141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp 16

562 CAAGTACCACTATATCAGGAACATGTTCAACCAATGGCAAGTGGGTGAACACAGGAC 62

161 ~~AsnGluThrGlyTyrTrpAsnValIleVal~~AsnSerProGluLysGlyAla 18

58

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

D6 682 GAGACATGGTCTGATTCCCTACGALGIICCAATAAAGTGACTTCGTGC

201 AAGIupheglYAlAGIuphelysAnIlegIutrasnIyItnIarGlepeleuIyI 22

Db 742 GCTGAATTTGGAGCAGAGTCAAGAACATAGAAACCACTATATACAAGAATATTTCTTAC 80

221 SERGIYGLUPROTHRTRYLEUGLYASNGLUThrSerValPheGlyPROThrgLYASnLYs 24

\_\_\_\_\_

QY	ProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyrGlu	61
Db	214 CCGNGTATGAGAGGAGCAVGAAYATGARGTNTTGMGYTTCARGCNCNGTNTGGAR	273
QY	62 PheIysTyrGlyAspLeuLeuGlyHisLeuIysIleMetHisAspAlaIleGlyIysPheArg	81
Db	274 TTYAATATAGGAGCAVYNTYNTGNCAYYNTAARAHATCAAGCAVGCNATHGGNTTGMGN	333
QY	82 SerThrLeuThrGlyIysAsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsn	101
Db	334 MSNACAYTNACGNGAARAAATYTAACNATGGARTGGTATGARTTNTTYCARVTNGNAAV	393
QY	102 CysThrPheProHisIleLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAla	121
Db	394 TGTACNTTTCNCAVYTNMGNCNGARATGAGAGCCNTTYYGTGYAAVCAARGGNGCN	453
QY	122 AlaCysPhePheGluGlyIleAspAspValHisTyrIysGluAsnGlyThrLeuValGln	141
Db	454 GCNTGTTTYYTARGGNAITHGAYGYNCAVYTGGAARARAAVGGNACNTYNTGNCAR	513
QY	142 ValAlaThrTrileSerGlyAsnMetPheAsnGluMetAlaIysTyrValIysGlnAspAsn	161
Db	514 GINGCACACNATHMSGNGAAYATGTTTAAACATATGGCNAARCGGCTNARCARGAATAV	573
QY	162 GluThrGlyIleTyrTyrGluThrTyrPheAsnValIysAlaSerProGluIysGlyAlaGlu	181
Db	574 GARAACGNAATHATATAYAGARACNTGGAAAGTNAAGCMMSCNCGARARARGGNGCAR	633
QY	182 ThrTyrPheAspSerTyrAspCysSerIysPheValLeuArgThrPheAsnIysLeuAla	201
Db	634 ACNTGTTTGAAYMSNTAAGAYATGYMSNAATTTGMYTNGACNTTAAVAAATYNGCN	693
QY	202 GluPheGlyValGluPheIysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSer	221
Db	694 GARTTYGNGCNGARITTYAARAAVATGARACNAAATAYACMGNAHTTYYNTATYMSN	753
QY	222 GlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIysThr	241
Db	754 CGNARCCNACNATATYNTGNGAAVGAARACMWSNGTNTTYYGNCNACNGGAAVAAARACN	813
QY	242 LeuGlyLeuAlaIleIysValArgPheTyrTyrProPheIysProHisIleuProThrIysGlu	261
Db	814 YTNAGNTYNGCNAATHAARMGNTTATATATACNTTYAACCCNCAVYNTCCNACNAAARGAR	873
QY	262 PheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisIysGlnPheTyrLeu	281
Db	874 TTYVTYNTMNSYNTYNTNCARAHHTTGAAGCNGTNAITHGNCAYAARCAATTTTAYTNT	933
QY	282 PheTyrAsnPheGluTyrTyrPheLeuProMetIysPheProPheIleIysIleThrTyr	301
Db	934 TTYTAATAATTTTGARKATVGTGTTTNTMCCNATGAARTTCNTTATATAAATAACNTAY	993
QY	302 GluGluIleProLeuProIleArgAsnIysThrLeuSerGly	315
Db	994 GARGARATHCCNNTNCCNATMGNAATAAARACNNTYNSNGCN	1035
RESULT 7		
LOCUS	BC025487	2318 bp mRNA linear ROD 16-APR-2002
DEFINITION	Mus musculus ceroid-1:lipofuscinosis, neuronal 5, mRNA (cDNA clone IMAGE:5251891), partial cds.	
ACCESSION	BC025487	
VERSION	BC025487.1 GI:19343790	
KEYWORDS		
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (baes 1 to 2318)	
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klauer,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altchek,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,	





[illegible]

```

TITLE      JOURNAL
REFERENCE  JOURNAL
AUTHORS    Wright,D., Wright,R., Wu,J., Yakus,S., Yen,J., Yosh,L., Yosh,V.,
            Yu,F., Zhang,J., Zhou,T., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstein,G. and Gibbs,R.A.
JOURNAL    Direct Submission
           Unpublished
           2 (baaes 1 to 198524)

REFERENCE  JOURNAL
TITLE      JOURNAL
AUTHORS    Morley,K.C.
JOURNAL    Direct Submission
           Submitted (05-FEB-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
           3 (baaes 1 to 198524)
           Rat Genome Sequencing Consortium.
           Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
           On Oct 11, 2002 this sequence version replaced gi:21738217.
           The sequence in this assembly is a combination of BAC based reads
           and whole genome shotgun sequencing reads assembled using Atlas
           (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
           in the feature table below represents a scaffold in the Atlas
           assembly (a 'contig-scaffold'). Within each contig-scaffold,
           individual sequence contigs are ordered and oriented, and separated
           by sized gaps filled with Ns to the estimated size. The sequence
           may extend beyond the ends of the clone and there may be sequence
           contigs within a contig-scaffold that consist entirely of whole
           genome shotgun sequence reads. Both end sequences and whole genome
           shotgun sequence only contigs will be indicated in the feature
           table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GOLF
Center clone name: CH230-331D3

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 177481 bases at least Q40
Consensus quality: 179640 bases at least Q30
Consensus quality: 181141 bases at least Q20
Estimated insert size: 183954; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 198524: contig of 198524 bp in length.

FEATURES
source
1..198524
   /organism="Rattus norvegicus"
   /mol_type="genomic DNA"
   /db_xref="taxon:10116"
   /clone="CH230-331D3"
1..1183
   /note="wgs_end_extension
clone_end:T"
131403..131443
   /note="clone_boundary
clone_end:T"
site:Mbol
end_sequence:RXAPCIATY"

```





REFERENCE 2 (bases 1 to 227920)  
 AUTHORS Worley, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 227920)  
 REFERENCE Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On May 13, 2003 this sequence version replaced gi:22855855. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GIBR  
 Center clone name: CH230-13816  
 ----- Summary Statistics

Assembly program: Atlas 3.0;  
 Consensus quality: 214528 bases at least Q40  
 Consensus quality: 21733 bases at least Q30  
 Consensus quality: 219064 bases at least Q20  
 Estimated insert size: 223690; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
 (see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 223559: contig of 223559 bp in length  
 \* 223560 223659: gap of unknown length  
 \* 223660 224738: contig of 1079 bp in length  
 \* 224739 224838: gap of unknown length  
 \* 224839 226289: contig of 1451 bp in length  
 \* 226290 226389: gap of unknown length  
 \* 226390 227920: contig of 1531 bp in length.

FEATURES  
 source 1..227920  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-13806"  
 1..1145  
 /note="wgs contig"  
 222227..223559  
 /note="wgs contig"  
 misc\_feature  
 misc\_feature  
 misc\_feature  
 /note="wgs\_contig"

ORIGIN  
 Alignment Scores: 1.7e-124 Length: 227920  
 Pred. No.: 1339.00 Matches: 235  
 Score:

Percent Similarity: 89.35% Conservative: 25  
 Best Local Similarity: 80.76% Mismatches: 31  
 Query Match: 76.17% Indels: 0  
 DB: 2 Gaps: 0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AC107510 (1-227920)

QY 7 SerArgAGhiSTRProValProTyrIysArgPheAspPheArgProLysProLysProPro 26  
 Db 37327 TCTGGGCAACGCTGGCCCGTACCTCAACAGCGCTTCCTCCGTCGGAAGCAATCC 37386  
 QY 27 TyrGyGlnAlaLysTyrThrPheCyProthrhGlySerProLysProValMetGluGly 46  
 Db 37387 TCTCTGCAAGCCAAAGTAAATACCTTCCTCCCTACCTGCTCAACCTCCAGTTTGAAGAGC 37446  
 QY 47 AspAspPheIleGluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAsp 66  
 Db 37447 AATGACGTATGAACTTAAGCTTAAAGCTCAAGCCCGTTGGGAATTAAATATGAGAGC 37506  
 QY 67 LeuLeuGlyValLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGly 86  
 Db 37507 CTCTCGGACACTTCAAAATCATGACGATGCTGATTCAGATTCAGAGGTACGCTGACCGGC 37566  
 QY 87 LysAsnTyrThrMetGluTyrPyrGluLeuPheGlnLeuGlyAsnCyThrPheProHis 106  
 Db 37567 AAGAACTACAAATTAATGATGATGAGCTTTCCAGCTCGGCACTGTCATTTCCCCAC 37626  
 QY 107 LeuArgProGluMetCysAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGlu 126  
 Db 37627 CTTCGCGCGGAAGTGAACGCTCCGTTCTGTGTATACAGAGGCGACCTGCTTTTGA 37686  
 QY 127 GlyIleAspAspValHisTyrPylGluAsnGlyThrLeuValGlnValAlaThrIleSer 146  
 Db 37687 GGAATTTACGATTAACACCTGAAAGAAACGGGACCTGCTGCTGCTGACCACTATCC 37746  
 QY 147 GlyAsnMetCysPheAsnGlnMetAlaLysTyrPylValGluAspAsnGluThrGlyIleTyr 166  
 Db 37747 GGAAACACGTTTAAACAAAGTGGCCGAGTGGTGAAGCAGCAATGAGACTGGGATTAT 37806  
 QY 167 TyrGluThrTyrAsnValLysAlaSerProGluLysGlyAlaGluThrTyrPheAspSer 186  
 Db 37807 TACGAGCATGACAGCTCCGCGCACGCCAGAAAGGGCGGACATGTTGAATCC 37866  
 QY 187 TyrAspCysSerLysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGlu 206  
 Db 37867 TACGACTGTTGCAATTTTGTCTTAAGGACATATGAGAAATTTGCTGATTTGAAACGAA 37926  
 QY 207 PheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyr 226  
 Db 37927 TTCAAGAAAGTGAAGAACAAATCATATGAAATATTTCTTACAGCGGAGAAACGATTAC 37986  
 QY 227 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle 246  
 Db 37987 TTGGGAATGAAAGTATATTTTGGGCCCAAGAAAGAAAGCCCTTGCTTGGCATA 38046  
 QY 247 LysArgPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuSerLeu 266  
 Db 38047 AAAAATTTTACGCGCCCGTCMAACGGTATTCGTCAACCAAGATTTTCTGTAATTTC 38106  
 QY 267 LeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGlu 286  
 Db 38107 TTGAAATTTTGAACACAGTATATATGACAGAGAGTTCTACGTTTATTAACCTTGAAG 38166  
 QY 287 TyrTrpPheLeuProMetLysPheProPheIle 297  
 Db 38167 TATTGTTTCTACCTATGAACCCCTTGTCTC 38199

RESULT 10  
 AC001226 106988 bp DNA linear PRI 29-MAY-1997  
 LOCUS AC001226  
 DEFINITION Genomic sequence from Human 13, complete sequence.  
 AC001226  
 AC001226.1 GI:2133862

WORDS HTG.  
 HOMO sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 PERCENCE 1 (bases 1 to 106988)  
 AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,  
 Fasmann,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,  
 Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Frilipp,W.U.,  
 Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,  
 Margulis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,  
 Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,  
 Stowell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and  
 Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 PERCENCE 3 (bases 1 to 106988)  
 AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,  
 Fasmann,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,  
 Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Frilipp,W.U.,  
 Gage,D., Geraigery,K., Hagos,B., Jacotot,L., Lane,M., Mackenzie,J.,  
 Margulis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,  
 Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,  
 Stowell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and  
 Zody,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 COMMENT On May 29, 1997 this sequence version replaced gi:1932720.  
 The Staden databases, finishing information, and all  
 chromatographic files used in the assembly of this clone are  
 available from our anonymous ftp site.  
 All repeats were identified using RepeatMasker: Smtt, A.F.A. &  
 Green, P. (1996-1997) Washington.edu/RM/RepeatMasker.html.  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
 FEATURES  
 source  
 1. 106988  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /clone\_id="76N15"  
 /clone\_lib="unknown"  
 complement(113..200)  
 repeat\_region  
 /rpt\_family="THR1B"  
 complement(202..611)  
 repeat\_region  
 /rpt\_family="L1PA3"  
 complement(674..694)  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 complement(835..981)  
 repeat\_region  
 /rpt\_family="MIR2"  
 1632..1920  
 repeat\_region  
 /rpt\_family="AluJo"  
 1970..2271  
 repeat\_region  
 /rpt\_family="AluSp"  
 2779..2946  
 repeat\_region  
 /rpt\_family="FAM"  
 3552..3928  
 repeat\_region  
 /rpt\_family="AluJb"  
 4361..4640  
 repeat\_region  
 /rpt\_family="AluSg"  
 5253..5385  
 repeat\_region  
 /rpt\_family="MLT1F"  
 5914..6106  
 repeat\_region  
 /rpt\_family="MLT1F"  
 complement(6341..6435)  
 repeat\_region

repeat\_region  
 /rpt\_family="MER34"  
 complement(6838..7140)  
 repeat\_region  
 /rpt\_family="AluSx"  
 7394..7455  
 repeat\_region  
 /rpt\_family="MER5A"  
 7464..7627  
 repeat\_region  
 /rpt\_family="MER5B"  
 complement(7624..7792)  
 repeat\_region  
 /rpt\_family="MER5B"  
 complement(8363..8391)  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 8480..8639  
 repeat\_region  
 /rpt\_family="MER5B"  
 complement(9199..9286)  
 repeat\_region  
 /rpt\_family="(TA)n"  
 9340..9644  
 repeat\_region  
 /rpt\_family="AluJo"  
 9880..9946  
 repeat\_region  
 /rpt\_family="MIR2"  
 10208..10787  
 repeat\_region  
 /rpt\_family="L1MB3"  
 10984..11297  
 repeat\_region  
 /rpt\_family="AluSg"  
 11311..11470  
 repeat\_region  
 /rpt\_family="L1PA8"  
 complement(11471..11495)  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 complement(12394..12440)  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 13782..13921  
 repeat\_region  
 /rpt\_family="MIR"  
 14035..14316  
 repeat\_region  
 /rpt\_family="AluSx"  
 complement(15018..15046)  
 repeat\_region  
 /rpt\_family="AT\_rich"  
 15080..15373  
 repeat\_region  
 /rpt\_family="AluJb"  
 15568..15881  
 repeat\_region  
 /rpt\_family="AluJo"  
 complement(18038..18071)  
 repeat\_region  
 /rpt\_family="(CAT)n"  
 complement(20389..20494)  
 repeat\_region  
 /rpt\_family="FAM"  
 21238..21534  
 repeat\_region  
 /rpt\_family="AluSx"  
 complement(23331..23353)  
 repeat\_region  
 /rpt\_family="GC\_rich"  
 complement(23546..23578)  
 repeat\_region  
 /rpt\_family="GC\_rich"  
 24565..24707  
 repeat\_region  
 /rpt\_family="MIR"  
 26013..26047  
 repeat\_region  
 /rpt\_family="MIR2"  
 complement(27717..28532)  
 repeat\_region  
 /rpt\_family="L1PB1"  
 complement(28553..28854)  
 repeat\_region  
 /rpt\_family="AluSg"  
 complement(28860..29076)  
 repeat\_region  
 /rpt\_family="L1"  
 complement(29365..29561)  
 repeat\_region  
 /rpt\_family="MIR"  
 complement(30050..30147)  
 repeat\_region  
 /rpt\_family="MIR"  
 31487..31812  
 repeat\_region  
 /rpt\_family="AluJo"  
 complement(33629..33732)  
 repeat\_region  
 /rpt\_family="MIR"  
 34434..34729  
 repeat\_region  
 /rpt\_family="AluSx"  
 35070..35361  
 repeat\_region  
 /rpt\_family="AluSx"  
 complement(36437..36461)  
 repeat\_region  
 /rpt\_family="AT\_rich"

```

repeat_region complement (39967, .40261)
/rpt_family="AlusX"
repeat_region complement (40741, .40915)
/rpt_family="AlusG"
repeat_region 40963, .41663
/rpt_family="L1PA2"
repeat_region 42042, .42352
/rpt_family="MLTIE"
repeat_region 42467, .42771
/rpt_family="AlusG"
repeat_region 42772, .42848
/rpt_family="L1MB8"
repeat_region 42849, .43151
/rpt_family="AlusB"
repeat_region 43152, .43318
/rpt_family="L1MA10"
repeat_region 43253, .43427
/rpt_family="L1MB8"
repeat_region 43429, .43672
/rpt_family="L1MB8"
repeat_region complement (44188, .44481)
/rpt_family="AlusC"
repeat_region 44544, .45246
/rpt_family="LTR8"
repeat_region complement (46999, .47032)
/rpt_family="AT_rich"
repeat_region complement (47411, .47431)
/rpt_family="AT_rich"
repeat_region complement (47612, .47684)
/rpt_family="CATAn"
repeat_region complement (47946, .48239)
/rpt_family="AlusB"
repeat_region complement (48827, .49872)
/rpt_family="AT_rich"
repeat_region complement (50160, .50186)
/rpt_family="AT_rich"
repeat_region 50952, .51121
/rpt_family="MERBA"
repeat_region complement (52140, .52284)
/rpt_family="AlusB"
repeat_region 53772, .53974
/rpt_family="MERB"
repeat_region 54225, .54348
/rpt_family="MIR"
repeat_region complement (54533, .54579)
/rpt_family="(TGA)n"
repeat_region complement (54580, .54736)
/rpt_family="AlusB"
repeat_region complement (54755, .54782)
/rpt_family="(CA)n"

```

```

Alignment Scores:
Pred. No.: 1,51e-82 Length: 106988
Score: 925.50 Matches: 173
Percent Similarity: 89.69% Conservative: 1
Best Local Similarity: 89.18% Mismatches: 1
Query Match: 52.65% Indels: 19
Gaps: 1

```

US-10-010-050a-2\_COPY\_31\_346 (1-316) x ACC001226 (1-106988)

```

QY 123 CysPhepHeGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuValGlnVal 142
Db 31851 TGTGTTTAA----- 31862
QY 143 AlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTTPValIleGlnAspAsnGlu 162
Db 31863 -----CTAGGAACATGTTCAACCAATGCGCAAGTGGGTAAACAGACATGAA 31913
QY 163 ThrGlyIleTyrTyrGlnThrTrpAsnValLysAlaSerProGluLysGlyValIleGluThr 182
Db 31914 ACGGAAATTTATTAAGACATGAGATGTAAAGCCAGCCGAAAGGGGGCAGAGACA 31973

```

```

QY 183 TrpPheAspSerTyrAspCysSerLysPheValLeuArgTrpPheAsnLysLeuAlaGlu 202
Db 31974 TGGTTGATTCCTACGACTGTTCCAAATTGTGTTAGACCTTTACAAAGTTGGCTGAA 32033
QY 203 PheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGly 222
Db 32034 TTGGACGACAGCTTCAGAACATAGAACCAACATATACAGAAATTTCTTTACAGTGA 32093
QY 223 GluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeu 242
Db 32094 GAACCTACTTATCTGGGAAATGAAACATCTGTTTGGGCCAACAGGAAACAAAGCTCTT 32153
QY 243 GlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLysProThrLysGluPhe 262
Db 32154 GGTTCAGCATTAAGAAAGATTATTAACCTTCACCAACATTTGCCAATTAAGAAATTT 32213
QY 263 LeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeuPhe 282
Db 32214 CTGTTGAGCTCTTCACAAATTTTGTATGACAGATGTCGCAACAAACAGTTCTATTTGTTT 32273
QY 283 TyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheIleLysIleThrTyrGlu 302
Db 32274 TATAATTTTGAATATGTTGTTTACCTATGAAATTCCTTTATTAATAATTAACATATGAA 32333
QY 303 GluIleProLeuProIleAlaArgAsnLysThrLeuSerGlyLeu 316
Db 32334 GAAATCCCTTACCTATTCAGAAACAAACACTCTCTGTTTA 32375

```

RESULT 11  
AL136440  
LOCUS  
DEFINITION  
Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.  
ACCESSION  
AL136440.2 GI:10039473  
VERSION  
HTG; HTGS PHASE1; HTGS\_CANCELED.  
KEYWORDS  
Homo sapiens (human)  
SOURCE  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT  
Burton J.  
Direct Submission  
Submitted (10-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
request: clonerequest@sanger.ac.uk  
On Sep 8, 2000 this sequence version replaced gi:6982057.  
-----  
Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquerry@sanger.ac.uk  
-----  
Project Information  
Center project name: B185124  
-----  
Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; 108752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 1642085 bases at least Q40  
Consensus quality: 164521 bases at least Q30  
Consensus quality: 166040 bases at least Q20  
Insert size: 168062; sum-of-contigs  
Insert size: 164996; 1.7% error; agarose-fp  
Quality coverage: 4.08x in Q20 bases; sum-of-contigs Quality  
coverage: 4.16x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently consists of 14 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES	Location/Qualifiers
source	1..169362 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="13" /clone RP11-185124" /clone lib="RPCI-11.1"
misc_feature	1..3470 /note="assembly_fragment:00318 clone_end:T7 vector_side:left"
misc_feature	3501..16444 /note="assembly_fragment:00726 fragment_chain:1"
misc_feature	16345..19514 /note="assembly_fragment:00157 fragment_chain:1"
misc_feature	19615..22241 /note="assembly_fragment:00846 fragment_chain:1"
misc_feature	22342..25526 /note="assembly_fragment:00271 fragment_chain:1"
misc_feature	25627..33070 /note="assembly_fragment:01555 fragment_chain:1"
misc_feature	33111..67620 /note="assembly_fragment:00768" 67721..93413 /note="assembly_fragment:00828"
misc_feature	93514..114081 /note="assembly_fragment:01026"
misc_feature	114182..126206 /note="assembly_fragment:01600 fragment_chain:2"
misc_feature	126307..144583 /note="assembly_fragment:01198 fragment_chain:2"
misc_feature	144684..147136 /note="assembly_fragment:01132 fragment_chain:2"
misc_feature	147237..164925 /note="assembly_fragment:01005 fragment_chain:2"
misc_feature	165026..169362

**JOURNAL**  
Submitted (12-JUN-2001) sanger centre, hinxton, cambridge, UK. E-mail enquiries: humquerry@sanger.ac.uk  
CB10 1SA, UK. sanger@cam.ac.uk

On Aug 22, 2000 this sequence version replaced gi:8894433.

# Genome Center

Center: Sanger Centre

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquerry@sanger.ac.uk](mailto:humquerry@sanger.ac.uk)

Project Information

Center project name: b4453N22

## Summary Statistics

Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 220597 bases at least Q40

Consensus quality: 234299 bases at least Q30

Consensus quality: 241103 bases at least Q20

Insert size: 245867; sum-of-contigs

Insert size: 190986; agarose-fp

Quality coverage: 2.52x in Q20 bases; sum-of-contigs Quality coverage: 3.92x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently consists of 53 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 2130: contig of 2130 bp in length
* 2131 2230: gap of 100 bp
* 2231 5689: contig of 3459 bp in length
* 5689 5789: gap of 100 bp
* 5789 10045: contig of 4257 bp in length
* 10045 10146: gap of 100 bp
* 10146 12624: contig of 2478 bp in length
* 12624 12724: gap of 100 bp
* 12724 20993: contig of 8269 bp in length
* 20993 21093: gap of 100 bp
* 21093 28935: contig of 7843 bp in length
* 28935 29037: gap of 100 bp
* 29037 32230: contig of 3194 bp in length
* 32230 32331: gap of 100 bp
* 32331 35153: contig of 2823 bp in length
* 35153 35154: gap of 100 bp
* 35154 37649: contig of 2396 bp in length
* 37649 37650: gap of 100 bp
* 37650 40245: contig of 2496 bp in length
* 40245 40346: gap of 100 bp
* 40346 48420: contig of 8075 bp in length
* 48420 48521: gap of 100 bp
* 48521 52169: contig of 3649 bp in length
* 52169 52170: gap of 100 bp
* 52170 56443: contig of 4174 bp in length
* 56443 56544: gap of 100 bp
* 56544 63088: contig of 6545 bp in length
* 63088 63089: gap of 100 bp
* 63089 65281: contig of 2093 bp in length
* 65281 65381: gap of 100 bp
* 65381 67940: contig of 2559 bp in length
* 67940 67941: gap of 100 bp
* 67941 75029: contig of 6989 bp in length
* 75029 75030: gap of 100 bp
* 75030 85567: contig of 10438 bp in length
* 85567 85568: gap of 100 bp
* 85568 89150: contig of 3483 bp in length
* 89150 89250: gap of 100 bp
* 89250 93380: contig of 4130 bp in length
* 93380 93480: gap of 100 bp
* 93480 95619: contig of 2139 bp in length
* 95619 95719: gap of 100 bp
* 95719 100582: contig of 4873 bp in length
* 100582 100593: gap of 100 bp
* 100593 103959: contig of 3267 bp in length

```

## FEATURES

### source

```

* 103960 104059: gap of 100 bp
* 104060 110871: contig of 6812 bp in length
* 110872 110971: gap of 100 bp
* 110972 115555: contig of 4584 bp in length
* 115556 115557: gap of 100 bp
* 115558 118593: contig of 2938 bp in length
* 118594 118595: gap of 100 bp
* 118596 123530: contig of 4837 bp in length
* 123531 123630: gap of 100 bp
* 123631 127087: contig of 3457 bp in length
* 127088 127187: gap of 100 bp
* 127188 129684: contig of 2497 bp in length
* 129685 129784: gap of 100 bp
* 129785 132332: contig of 2348 bp in length
* 132333 132334: gap of 100 bp
* 132335 143792: contig of 11560 bp in length
* 143793 143892: gap of 100 bp
* 143893 148766: contig of 4874 bp in length
* 148767 148866: gap of 100 bp
* 148867 152358: contig of 3392 bp in length
* 152359 152359: gap of 100 bp
* 152360 156922: contig of 4564 bp in length
* 156923 157022: gap of 100 bp
* 157023 160153: contig of 3131 bp in length
* 160154 160253: gap of 100 bp
* 160254 164334: contig of 4081 bp in length
* 164335 164335: gap of 100 bp
* 164336 166000: contig of 2166 bp in length
* 166001 166700: gap of 100 bp
* 166701 171896: contig of 5196 bp in length
* 171897 171996: gap of 100 bp
* 171997 184915: contig of 12919 bp in length
* 184916 185015: gap of 100 bp
* 185016 187238: contig of 2223 bp in length
* 187239 187338: gap of 100 bp
* 187339 189602: contig of 2264 bp in length
* 189603 189702: gap of 100 bp
* 189703 194563: contig of 4761 bp in length
* 194564 194664: gap of 100 bp
* 194665 197682: contig of 3119 bp in length
* 197683 197782: gap of 100 bp
* 197783 201788: contig of 4006 bp in length
* 201789 201888: gap of 100 bp
* 201889 204060: contig of 2172 bp in length
* 204061 204160: gap of 100 bp
* 204161 206237: contig of 2077 bp in length
* 206238 206337: gap of 100 bp
* 206338 210376: contig of 4039 bp in length
* 210377 210476: gap of 100 bp
* 210477 225245: contig of 14769 bp in length
* 225246 225345: gap of 100 bp
* 225346 230027: contig of 4682 bp in length
* 230028 230127: gap of 100 bp
* 230128 232610: contig of 2483 bp in length
* 232611 232710: gap of 100 bp
* 232711 234942: contig of 2232 bp in length
* 234943 235043: gap of 100 bp
* 235044 240958: contig of 5916 bp in length
* 240959 241058: gap of 100 bp
* 241059 251187: contig of 10129 bp in length.

```

### Location/Qualifiers

```

1. 251187
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="13"
  /clone_id="RP11-453N22"
  /clone_lib="RPC1-11.2"
  1. 2130
  /note="assembly_fragment:00469
  fragment_chain:1"
  2231. 5689
  /note="assembly_fragment:01219

```

### misc\_feature

### misc\_feature

QY	243	GlyleuaiaallieysaGphneTYrrPrropheyProhiisleuProthrlvsGIupe	262
Dd	193804	GGTTTAGCCATAAAAGATTTTATTACCCTCCCAACACACTTGCCACTTAAGAATTT	193745
QY	263	LenuSertLeuileGnillePheAsPaIavaIIeValHislsyGlnPhetyrLsuphe	282
Dd	193744	CAGGTGAGCTCTTGCAAAATTTTGATGCAGTATTTGCACAACAGTTCMAATTTGTTT	193685
QY	283	TyrasnpheGLUTyrTrpPheLeuProWetLyaspheProhelleyslleTHrTYrGlu	302
Dd	193684	TATATTTGAATATTTGGTTTTTTACCATGAATTCCTCTTTAATTAAAAATACATATGAA	193625
QY	303	GiulleProleuproillearqasmlyThleuserglyden	316
Dd	193624	GAATCCTTTACTATCAGAAACAAAACACTCTCGTTTA	193583
RESULT 13			
BD125187			
LOCUS	BD125187	697 bp	DNA linear PAT 18-SEP-2002
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.		
ACCESSION	BD125187		
VERSION	BD125187.1 GI:23220132		
KEYWORDS	JP 2002017375-A/618, JP 2002017375-A/618,		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 697) Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Negai,K., Kojima,S., Otsuki,T. and Koga,H. TITLE Primer for synthesizing full-length cDNA and use thereof JOURNAL Patient: JP 2002017375-A 618 22-JAN-2002; HELIIX RESEARCH INSTITUTE OS Homo sapiens (human) PN JP 2002017375-A/618 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO PI ISHII, PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI, HISASHI KOGA PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC 10' C1221/02,C1201/68//C12P2/08,G06F17/30,C12N15/00,C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key Location/Qualifiers FT source 1..697 FT location/Qualifiers /organism='Homo sapiens (human)'. 1..697 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"		
FEATURES			
SOURCE			
ALIGNMENT SCORES:			
Pred. No.:	2.38e-82	Length:	697
Score:	896.00	Matches:	170
Percent Similarity:	92.51%	Conservative:	3
Best Local Similarity:	90.91%	Mismatches:	8
Query Match:	50.97%	Indels:	2
DB:	6	Gaps:	2
US-10-010-05DA-2_COPY_31_346 (1-316) x BD125187 (1-697)			
QY	1	ArgvalsergylileProSerArgArgHisETrpProValProTyrylysargPheapphe	20
Dd	142	GGGGGCTGGGGGAGATCCCTCCCGGCGCACATGGCCGCGGCCCGCGAACGCGCTTGACTTC	201



```

21 ArgProlylsProAspProTyrCysGlnAlaIleValSerThrPheCysProThrGlySerPro 40
202 CGTCCAAAACCTGATCCTTATGTCAGCTAAGTACTTCTGTCACACTGGCTCACT 261
41 IleProValMetGlnGlyAspAspAspIleGlnValPheArgLeuGlnAlaProValTyrP 60
262 ATCCAGATTATGAGAGGATGATGACACTGAGATTTCATTTCATTCAGACCCCACTATGG 321
61 GluPheValTyrGlyAspLeuLeuGlyIleValSerIleMetHisAspAlaIleGlyPhe 80
322 GAATTAAATATGAGACCTCTCGGACACTTGAATAATATGACAGATGATGATGATTC 381
81 ArgSerThrIleuThrGlyValSerThrMetGluTyrPylGluLeuPheGlnLeuGly 100
382 AGAAGTACTTAACTGCGCAAGACATCAACATGATGATGATGATGATGATGATGATGATG 441
101 AsnCysThrPheProHisIleuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
442 AACTGTCATTTCCCATCTCCGACCTGGAATGGAATGCCCTTTCGGTGTATCAAGGC 501
121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTyrPylGlnValSerIleuVal 140
502 GCGCTGCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 561
140 IginValAlaThrIleSer-GlyAsnMetPheAsnGlnMetAla-LysTyrValIleGln 159
562 TCAGTAGCACTATATCAAGAACATGTTCAACCAATGCGCAAAAGCGGTGAACAA 621
160 ---AspAsnGluThrGlyIleTyrGlyIleThrIleAsnValIleAsnSerProGly 173
622 GGACATNGAACAGAGAAATTATATGACATGGAATTTAAAGCACCANNA 681
178 s---GlyAlaGlu 181
682 AAANGGGGCAAAA 694
b
RESULT 14
OCUS BD126453 697 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD126453
VERSION BD126453.1 GI:2221398
KEYWORDS UP 2002017375-A/1884.
SOURCE Homo sapiens (human).
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
Koga, H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 1884 22-JAN-2002;
HELIX RESEARCH INSTITUTE
COMMENT
OS Homo sapiens (human)
PN UP 2002017375-A/1884
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA, HISASHI KOGA
PI TETSUO OTA, HISASHI KOGA
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/
10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N5/00, C12N5/00
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1. 697
PT /organism='Homo sapiens (human)',
location/Qualifiers
1. 697
FEATURES
source

```

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 2,386-82 Length: 697
Score: 896.00 Matches: 170
Percent Similarity: 92.51% Conservative: 3
Best Local Similarity: 90.91% Mismatches: 8
Query Match: 50.97% Indels: 6
DB: Gaps: 2
US-10-010-050A-2_copy_31_346 (1-316) x BD126453 (1-697)
QY 1 ArgValSerGlyIleProSerArgArgHisTyrProValProTyrIleAspPheAspPhe 20
Db 142 CGGGTCTCGGCAATCCCTCCCGGCGCACACGCGCGGCTGGAAGCGCTTGACATTC 201
QY 21 ArgProlylsProAspProTyrCysGlnAlaIleValSerThrPheCysProThrGlySerPro 40
Db 202 CGTCCAAAACCTGATCCTTATGTCAGCTAAGTACTTCTGTCACACTGGCTCACT 261
QY 41 IleProValMetGlnGlyAspAspAspIleGlnValPheArgLeuGlnAlaProValTyrP 60
Db 262 ATCCAGATTATGAGAGGATGATGACACTGAGATTTCATTTCATTCAGACCCCACTATGG 321
QY 61 GluPheValTyrGlyAspLeuLeuGlyIleValSerIleMetHisAspAlaIleGlyPhe 80
Db 61 TCAGTAGCACTATATCAAGAACATGTTCAACCAATGCGCAAAAGCGGTGAACAA 621
QY 160 ---AspAsnGluThrGlyIleTyrGlyIleThrIleAsnValIleAsnSerProGly 173
Db 322 GAATTAAATATGAGACCTCTCGGACACTTGAATAATATGACAGATGATGATGATGATG 381
QY 81 ArgSerThrIleuThrGlyValSerThrMetGluTyrPylGluLeuPheGlnLeuGly 100
Db 382 AGAAGTACTTAACTGCGCAAGACATCAACATGATGATGATGATGATGATGATGATGATG 441
QY 101 AsnCysThrPheProHisIleuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
Db 442 AACTGTCATTTCCCATCTCCGACCTGGAATGGAATGCCCTTTCGGTGTATCAAGGC 501
QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTyrPylGlnValSerIleuVal 140
Db 502 GCGCTGCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGATGATGATGATG 561
QY 140 IginValAlaThrIleSer-GlyAsnMetPheAsnGlnMetAla-LysTyrValIleGln 159
Db 562 TCAGTAGCACTATATCAAGAACATGTTCAACCAATGCGCAAAAGCGGTGAACAA 621
QY 160 ---AspAsnGluThrGlyIleTyrGlyIleThrIleAsnValIleAsnSerProGly 173
Db 622 GGACATNGAACAGAGAAATTATATGACATGGAATTTAAAGCACCANNA 681
QY 178 s---GlyAlaGlu 181
Db 682 AAANGGGGCAAAA 694
b
RESULT 15
AE014175 0
WPCOMMENT
Sequence split into 4 fragments LOCUS AE014175 Accession AE014175
Fragment Name Begin End
AE014175_1 1 110000
AE014175_2 100001 210000
AE014175_3 200001 310000
AE014175_4 300001 404829
LOCUS AE014175 404829 bp DNA linear ROD 06-AUG-2002
DEFINITION Mus musculus p16a1d deletion region section 3 of 11 of the
complete sequence.
ACCESSION AE014175 AE0135600
VERSION AE014175.1 GI:22128029
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 404829)  
 Peterson, K., King, B., Hagg-Greenberg, A., Rolz, J., Bult, C. and  
 O'Brien, T.  
 Functional and comparative genomic analysis of the piebald deletion  
 region of mouse chromosome 14  
 Genomics 80 (2), 172 (2002)  
 JOURNAL 22150871  
 MEDLINE 12160731  
 PUBMED 12160731  
 2 (bases 1 to 404829)  
 Adams, M. and Mural, R.  
 Direct Submission  
 Submitted (31-MAY-2002) Celera Genomics, 45 West Gude Drive,  
 Rockville, MD 20850, USA  
 The strings of n's in this record represent gaps between contigs,  
 and the length of each string corresponds to the length of the gap  
 except where there are 20 n's. A string of 20 n's represents a gap  
 of unknown length.

FEATURES  
 source  
 Location/Qualifiers  
 1. 404829  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="14"  
 /map="M164-M165"  
 <38137..>47216  
 /gene="Trg1"  
 <38137..>47216  
 /gene="Trg1"  
 /product="Immune responsive gene 1"  
 /note="partial transcript sequence derived from alignment  
 to GenBank Accession Number L38281"  
 <61259..>68607  
 /gene="Cln5"  
 join(<61259..61421,62721..62886,64274..64499,  
 66916..>68607)  
 /gene="Cln5"  
 /product="ceroid-lipofuscinosis, neuronal 5"  
 /note="partial transcript derived from alignment to  
 tentative consensus sequence TC227789, from TIGR Mouse  
 Gene Index (v. 6.0)."  
 complement(<73858..>86476)  
 /gene="Fbl3a"  
 complement(join(<73858..74501,80293..80464,83295..83417,  
 86126..>86476))  
 /gene="Fbl3a"  
 /product="leucine-rich repeat-containing F-box protein  
 Fbl3a"  
 complement(join(73858..74501,80293..80464,83295..83417,  
 86126..>86476))  
 /codon\_start=1  
 /gene="Fbl3a"  
 /product="leucine-rich repeat-containing F-box protein  
 Fbl3a"  
 /protein\_id="AAM92567.1"  
 /db\_xref="GI:22128030"  
 /translation="RMKRGPRSDSDSAREGTAEKPRPTTQERSOPCDMGNLLQDI  
 VHVFRYLPLIDRAHSQVCRNMQVFHMPDLRCRPEFLNCPATSYLAKTHBELIKO  
 IIRHSNHLQVYSPKVDSSKESAACDILSQVNCSLKTLGLISTARSPMDPKSH  
 FIALTFVFNSSKSLSLKIDTPVDDPSLKLVANNSDTLKLRSSCPHVSFAGIL  
 CVADOCHEBELALNYHLSDBLLAISSEKHYLEHLIDVSENPQGHFTTICKS  
 SMRAFIKHSPKLVLVNVPFLYEREPDFRFEYIPATHLYGRSVSKDVGAVGTICPS  
 LVELVVCANGRLPDLBELIRIARCKNLASIGICEVSCSAFVEFYKMGGRSLGCLS  
 IMEEVLIPOQKYSLEQIHMEVSKHLGRVWFPDMPTW"  
 complement(<106136..>339446)  
 /gene="Pnr1"  
 complement(<106136..>339446)  
 /gene="Pnr1"  
 /product="Pam/hw/rpm-1"  
 /note="PAM, protein associated with Myc; partial  
 transcript derived from alignment of Homo sapiens NCBI  
 RefSeq sequence NM\_015057"

ORIGIN

Alignment Scores:  
 Pred. No.: 1.7e-62 Length: 110000  
 Score: 730.00 Matches: 127  
 Percent Similarity: 87.65% Conservative: 22  
 Best Local Similarity: 74.71% Mismatches: 21  
 Query Match: 41.52% Indels: 0  
 DB: 10 Gaps: 0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AE014175\_0 (1-110000)

QY	147	GIYASMECHPEASNGIMETALALYETPVALYSGLSASPAANGIUNRGLYLETYR	166
DB	66915	GGAACACATTATACAAAGTGGCCGATGGTGAACAGACATGAACCTGGATTAT	66974
QY	167	TYRGLUHTTRPASNVALYSAJASERPROGLULYSGLYALAGLUHTTRPHEASPSER	186
DB	66975	TATAGACATGAGACGGCCGACGGCCGACGACAGACAGAGGGCCGACAGCTGGTTCAGATCC	67034
QY	187	TYRASPCTSERLYSPHEVALLEUATGTRPHEASNLYSLEUALAGIUPHEGLYALAGLU	206
DB	67035	TACGACTGTTGCAATTTTGCTTAAGACATVAGAAATGGCTGAATTTGGAACAGAA	67094
QY	207	PHELYASNLIEGLUHTRASNLYTRARGTLEPHELEUTYRSEGLYGLUPROTHRTRY	226
DB	67095	TTCAAGAGATGAAACAACTATACAAATATTTCTTACATGGAGACCTATTAC	67154
QY	227	LEUGLYASNGIUNRSEVALPHEGLYPROTHRGLYASNLYSTRLEUGLYLEUALALE	246
DB	67155	CTGGGAATGAAACATCTATTGTTGGCCCAAGAAACAAAGACTCTTGGCCATTA	67214
QY	247	LYSARGPHELYTRYRPROPHELYPROHISLEUPROTHRLYSGIUPHELEUSENLEU	266
DB	67215	AAAAAATTTTATGGCCCTTCAACCGATTTGTCACCAAGATTTCTGATGAATTC	67274
QY	267	LEUGNILEPHEASPALAVALLIYVALHISLYSGINPHELYLEUPHETRYASNPHEGLU	286
DB	67275	TTGAAATTTTGTATACGTGATTTATACAGACAGCTCTACTGTTTATTAACCTTGAG	67334
QY	287	TYRTPPHELEUPROMELIYSPHEPROPHEILEYSLIETHTYRGLUGLULEPROLEU	306
DB	67335	TATIGTTTCAACCAATGAAACCCCTTGTGCAAAATACATACGAAAGAACCCCGTTA	67394
QY	307	PROLEAARGASNLYTRLEUSENGLYLEU	316
DB	67395	CCTACCCGACATACCAACATTTACCGACTTG	67424

Search completed: April 25, 2004, 03:23:00  
 Job time : 6261.28 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame\_plus\_p2n model

1 on: April 24, 2004, 22:59:32 ; Search time 529.278 Seconds  
(without alignments)  
2292.945 Million cell updates/sec

file: US-10-010-050A-2\_COPY\_1\_28  
effect score: 154  
sequence: 1 MRRGAGAGRAGSACWALMLMLAVVPG 28

oring table: BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

arched: 3470272 segs, 2167151695 residues  
tal number of hits satisfying chosen parameters: 6340544

num DB seq length: 0  
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh  
=/cgn2\_1/USPTO\_SPOOL/US10010050/runat\_22042004\_113204\_27549/app\_query\_fasta\_1.1372  
B=genemb1 -QMT=fastap -SUFFIX=xge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
NITS=bits -START=1 -END=-1 -MATRIX=BIOSUM62 -TRANS=human40.cdi -LIST=45  
LOCALIGN=200 -THR SCORE=pcpt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
JUTPUT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
USER=US10010050 @CGN 1.1 7433 @runat\_22042004\_113204\_27549 -NCPU=6 -ICPU=3  
NO MAP -LARGEOBJECTY -NEG SCORES=0 -WAT -DSPBLOCK=100 -LONGLOG  
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
XGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_inv:\*  
32: em\_hg\_other:\*  
33: em\_hg\_mus:\*  
34: em\_hg\_pln:\*  
35: em\_hg\_rnd:\*  
36: em\_hg\_mem:\*  
37: em\_hg\_vit:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	154	100.0	BD125187	BD125187 Primer fo
2	154	100.0	BD126453	BD126453 Primer fo
3	154	100.0	BD073402	BD073402 Secreted
4	154	100.0	BD136338	BD136338 95 human
5	154	100.0	BD127858	BD127858 Primer fo
6	154	100.0	AK075109	AK075109 Homo sapi
7	154	100.0	AF068227	AF068227 Homo sapi
8	154	100.0	AC001226	AC001226 Genomic s
9	154	100.0	AL136440	AL136440 Homo sapi
10	154	100.0	AL358875	AL358875 Homo sapi
11	146	94.8	BD073413	BD073413 Secreted
12	133	47.4	MI067213	MI067213 Mesothelio
13	72	46.8	HSBA19D2	AL080248 Human DNA
14	71	46.1	AC079116	AC079116 Homo sapi
15	71	46.1	AC008177	AC008177 Homo sapi
16	71	46.1	AC007248	AC007248 Homo sapi
17	71	46.1	AC087615	AC087615 Homo sapi
18	70.5	45.8	AC074349	AC074349 Homo sapi
19	70.5	45.8	AF070499	AF070499 Sequence
20	70	44.8	AF075622	AF075622 Sequence
21	69	44.8	AK120197	AK120197 Oryza sat
22	69	44.8	AR385625	AF385625 Sus scrofa
23	69	44.8	AC117471	AC117471 Homo sapi
24	69	44.8	AL137007	AL137007 Human DNA
25	69	44.8	AP003906	AP003906 Oryza sat
26	69	44.8	AC004589	AC004589 Oryza sat
27	69	44.8	AC079984	AC079984 Homo sapi
28	69	44.8	AL161716	AL161716 Human DNA
29	69	44.8	AC110302	AC110302 Mus muscu
30	69	44.8	AP005442	AP005442 Oryza sat
31	69	44.8	SC0939115	AL393115 Streptomy
32	68	44.2	AC108168	AC108168 Homo sapi
33	68	44.2	AL606530	AL606530 Human DNA
34	68	44.2	AC091025	AC091025 Homo sapi
35	67.5	43.8	PM282B	AL684841 Penicilli
36	67.5	43.8	AL1596209	AL1596209 Mouse DNA
37	67.5	43.8	AC113407	AC113407 Homo sapi
38	67.5	43.8	AC021965	AC021965 Homo sapi
39	67.5	43.8	AC084044	AC084044 Mus muscu
40	67.5	43.8	CPWPCA	L29028 Chlamydomon
41	67	43.5	AL627434	AL627434 Human DNA
42	67	43.5	AC025259	AC025259 Homo sapi
43	67	43.5	AC116479	AC116479 Mus muscu
44	67	43.5	AL442068	AL442068 Homo sapi
45	66.5	43.2	HSA328178	AJ328178 Homo sapi

RESULT 1

## ALIGNMENTS

LOCUS	BD125187	697 bp	DNA	linear	PAT 18-SEP-2002
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BD125187				
VERSION	BD125187.1	GI:23220132			
KEYWORDS	JP 2002017375-A/618				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 697) Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.				
TITLE	Primer for synthesizing full-length cDNA and use thereof				
JOURNAL	Patent: JP 2002017375-A 618 22-JAN-2002; HELIX RESEARCH INSTITUTE				
COMMENT	OS Homo sapiens (human) PN JP 2002017375-A/618 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO PI YURI KAWAI,AI MAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUJI OTSUKI,HISASHI KOGA PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC 10',C12P21/02,C12J01/68/C12B21/08,G06F17/30,C12N15/00,C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH key Location/Qualifiers FT source 1..697 FI location/Qualifiers 1..697 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"				
FEATURES					
source					
ORIGIN					
Alignment Scores:					
Pred. No.:	1..14e-08	Length:	697		
Score:	154.00	Matches:	28		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatch:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		
US-10-010-050A-2_COPY 1_28 (1-28) x BD125187 (1-697)					
QY	1 Metatgatagtgialaagialaalaargialaargialaaseritpocyettipalaaleuallenu 20				
DB	52 ATGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGTGCTGCGGCGCTGCGCTG 111				
QY	21 leutrpaleuallavalprogly 28				
DB	112 CTTTGCTGCGCGCGCTGCGGCG 135				
RESULT 2					
BD126453					
LOCUS	BD126453	697 bp	DNA	linear	PAT 18-SEP-2002
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BD126453				
VERSION	BD126453.1	GI:23221398			
KEYWORDS	JP 2002017375-A/1884				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 697) Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and				

TITLE	Koga,H.
JOURNAL	Primer for synthesizing full-length cDNA and use thereof Patent: JP 2002017375-A 1984-22-JAN-2002; HELIX RESEARCH INSTITUTE
COMMENT	OS Homo sapiens (human) PN JP 2002017375-A/1884 PD 22-JAN-2002 PF 07-JUL-2000 JP 2000253172 PI TOSHIO OKA,ETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO PI IIHII, PT YUDI KAMAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PT SHINICHI KOIITA, PC TESTSUJI OTSUKI,HISASHI KOGA
FEATURES	C12N5/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ C12N15/02,C12Q1/68/C12P21/08,G06F17/30,C12N5/00 CC PC ' C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N5/00 CC Primer for synthesizing full-length cDNA and use thereof FH Key Location/Qualifiers FT source 1..697 /organism='Homo sapiens (human)'. FT location/Qualifiers 1..697 /organism='Homo sapiens' /mol_type='genomic DNA' /db_xref='taxon:9606'
ORIGIN	Alignment Scores: Pred. No.: 1,14e-08 Length: 697 Score: 154.00 Matches: 28 Percent Similarity: 100.00% Conservative: 0 Best Local Similarity: 100.00% Mismatches: 0 Query Match: 100.00% Indels: 0 DB: Gaps: 0
US-10-010-050A-2_COPY_1_28 (1-28) x BD126453 (1-697)	
Oy	1 MeCArGArGeTAlAcGAlAAAGAAGCGlVArGAAsErTPCySrTPPaLeuAlaLeu 20 
Dd	52 ATGGGGGGGGCCGGGGGGCGGCCCTTCGTGtGCTGGcCCCTGGGGCTG 111 
Oy	21 LeuTPleuaAlaValProGl 28 
Dd	112 CTTrGGCTCGGGGrGGrTTCCGGGG 135 
RESULT 3	
BD073402	1486 bp DNA linear PAT 27-AUG-2002
LOCUS	
DEFINITION	Secreted protein which human chromosome 13 encodes.
ACCESSION	BD073402
VERSION	BD073402.1 GI:22619005
KEYWORDS	JP 2001511345-A/I. Homo sapiens (human)
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1486) Sheppard,P.O. and Gilbertson,D.G. Secreted protein which human chromosome 13 encodes Patent: JP 2001511345-A I 14-AUG-2001;
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	ZYMOGENETICS INC
COMMENT	OS Homo sapiens (human) PN JP 2001511345-A/1 PD 14-AUG-2001 PF 24-JUL-1998 JP 2000504249 PR 24-JUL-1997 US 60/053613 PI PAUL O SHEPPARD,DIBRA G GILBERTSON PC C12N5/09,A61K38/00,C07K14/47,C07K16/18,C12N1/15, PC C12N1/19,
PC	C12N1/21,C12N5/10,C12P21/02,C12Q1/68,C12N15/00,A61K37/02,C12N5/ 00

CC Secreted protein which human chromosome 13 encodes FH Key  
 FT CDS Location/Qualifiers  
 (47) . (1084) .  
 source 1.1486  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

IGIN

Alignment Scores:  
 ad. No.: 2.22e-08 Length: 1486  
 Matches: 28  
 Score: 154.00  
 Percent Similarity: 100.00% Conservative: 0  
 Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 100.00% Gaps: 0

-10-010-050a-2\_copy\_1\_28 (1-28) x BD073402 (1-1486)

1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeu 20  
 47 ATGCGGCGGGCGCGCGCGCGCGCTTCGCGGAGCGCGCTTCGCGCGCGCGCTG 106  
 21 LeuTrpLeuAlaValAlaProGly 28  
 107 CTTGGCTCGCGGCGGTTCGCGGC 110

SUIT 4  
 136338 BD136338 1751 bp DNA linear PAT 18-SEP-2002  
 CUS 95 human secretory proteins.  
 DEFINITION BD136338  
 CESSION BD136338.1 GI:23231283  
 YWORDS JP 2002506627-A/25.  
 URCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1751)  
 Ruben, S.M., Ni, Y., Rosen, C.A., Yu, G.L., Young, P.E., Fen, P.,  
 Soppet, D.R., Wei, Y.F., Endress, G.A., Duan, R.D., Kyaw, H., Ehner, R.,  
 Lafleur, D.W., Olsen, H.S., Shi, Y. and Moore, F.A.  
 95 human secretory proteins  
 Patent: JP 2002506627-A 25 05-MAR-2002;  
 JOURNAL HUMAN GENOME SCIENCES INC  
 COMMENT OS Homo sapiens (human)  
 PN JP 2002506627-A/25  
 PD 05-MAR-2002  
 PF 18-MAR-1999 JP 2000536733  
 PR 19-MAR-1998 US 60/078566, 19-MAR-1998 US 60/078576 PR  
 19-MAR-1998 US 60/078573, 19-MAR-1998 US 60/078574 PR  
 19-MAR-1998 US 60/078579, 19-MAR-1998 US 60/078578 PR  
 19-MAR-1998 US 60/078581, 19-MAR-1998 US 60/078577 PR  
 19-MAR-1998 US 60/078563, 01-APR-1998 US 60/080314 PR  
 01-APR-1998 US 60/080312, 01-APR-1998 US 60/080313 PI STEVEN  
 M RUBEN, JIAN NI, CRAIG A ROSEN, GDUO  
 LIANG YU PAUL, E YOUNG,  
 PI PING FENG,  
 PI DANIEL R SOPPET, YING FEI WEI, GREGORY A ENDRESS, ROXANNE D DUAN,  
 PI HUA KYAW,  
 PI REINHARD EHNER, DAVID W LAFLEUR, HENRIK S OLSEN, YANGSU SHI, PAUL,  
 PI A MOORE  
 PC C12N1/15,  
 PC C12N1/15,  
 PC C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12O1/68, G01N3/53, G01N3/56,  
 C12N15/00, A61K37/02, C12N5/00  
 CC n equals a,t,g, or c  
 CC n equals a,t,g, or c  
 CC n equals a,t,g, or c  
 CC n equals a,t,g, or c

CC n equals a,t,g, or c  
 CC n equals a,t,g, or c  
 FH Key  
 FT source 1.1751  
 /organism="Homo sapiens (human)"  
 /db\_xref="taxon:9606"

FEATURES  
 source 1.1751  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

ORIGIN

Alignment Scores:  
 Pred. No.: 2.56e-08 Length: 1751  
 Matches: 28  
 Score: 154.00  
 Percent Similarity: 100.00% Conservative: 0  
 Mismatches: 0  
 Best Local Similarity: 100.00% Indels: 0  
 Query Match: 100.00% Gaps: 0

US-10-010-050a-2\_copy\_1\_28 (1-28) x BD136338 (1-1751)

QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeu 20  
 49 ATGCGGCGGGCGCGCGCGCGCGCTTCGCGGAGCGCGCTTCGCGCGCGCGCTG 108  
 DB 21 LeuTrpLeuAlaValAlaProGly 28  
 DB 109 CTTGGCTCGCGGCGGTTCGCGGC 132

RESULT 5  
 BD127858 2120 bp DNA linear PAT 18-SEP-2002  
 LOCUS BD127858  
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.  
 ACCESSION BD127858  
 VERSION BD127858.1 GI:23222803  
 KEYWORDS JP 2002017375-A/3289.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 2120)  
 Oca, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,  
 Wakamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and  
 Koga, H.  
 Primer for synthesizing full-length cDNA and use thereof  
 Patent: JP 2002017375-A 3289 22-JAN-2002;  
 JOURNAL HELIX RESEARCH INSTITUTE  
 COMMENT OS Homo sapiens (human)  
 PN JP 2002017375-A/3289  
 PD 22-JAN-2002  
 PF 07-JUL-2000 JP 2000253172  
 PI TOSHIO OCA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO  
 PI ISHII,  
 PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI  
 SHINICHI KOJIMA,  
 PI TETSUO OCA, HISASHI KOGA  
 PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/10,  
 C12P21/02, C12O1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC  
 Primer for synthesizing full-length cDNA and use thereof FH Key  
 FT CDS Location/Qualifiers  
 (16) . (1089) .  
 source 1.2120  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"

ORIGIN

Alignment Scores:  
 Pred. No.: 3.03e-08 Length: 2120



```

21 LeuTRpleu1aValValProgly 28
|||||
1536 CTTTGCTCGCGGTGTTCCGGGC 1559

SULT 8
001226 AC001226 106988 bp DNA linear PRI 29-MAY-1997
CUS Genomic sequence from Human 13, complete sequence.
FINITION AC001226
CESSION AC001226.1 GI:2133862
RSION HTG.
WORDS Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 106988)
Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H. and Lander,E.S.
TITLE Unpublished
JOURNAL 2 (bases 1 to 106988)
REFERENCE 2 (bases 1 to 106988)
AUTHORS Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Brown,K., Cooke,P., Daly,M.J., Forrest,C., Fritpp,W.J.,
Gage,D., Geradgery,K., Hages,B., Jacotot,L., Lane,M., Mackenzie,J.,
Margis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome
REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 106988)
Hawkins,T.L., Reeve,M.P., Christoffersen,A., Birren,B.W.,
Fasman,K.H., Lander,E.S., McKernan,K., Munro,C., Richardson,P.,
Barna,N., Chang,A., Cooke,P., Daly,M.J., Forrest,C., Fritpp,W.J.,
Gage,D., Geradgery,K., Hages,B., Jacotot,L., Lane,M., Mackenzie,J.,
Margis,N., McDermott,J., Moloney,N., Morrow,J., Nachman,A.,
Naylor,J., O'Connor,T., Peterson,K., Rollins,G., Spencer,J.,
Stilwell,J., Stone,C., Strickland,C., Sydney,K., Wilmer,F. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome
REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS On May 29, 1997 this sequence version replaced gi:1932720.
COMMENT The Staden databases, finishing information, and all
chromatographic files used in the assembly of this clone are
available from our anonymous ftp site.

All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
Location/Qualifiers
1. 106988
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="76N15"
/clone_lib="unknown"
/complement(113..200)
/rpt_family="THERB"
/complement(202..611)
/rpt_family="LIPB3"
/complement(674..694)
/rpt_family="AT_rich"
/complement(835..981)
/rpt_family="MIR2"
1632..1970
/rpt_family="AluB"
1970..2271
/rpt_family="MIR"
2779..2946
/rpt_family="MIR"

/rpt_family="FAM"
3552..3828
/rpt_family="AluB"
4361..4640
/rpt_family="AluSg"
5253..5385
/rpt_family="MLTIF"
5914..6106
/rpt_family="MLTIF"
complement(6341..6435)
/rpt_family="MER34"
complement(6838..7140)
/rpt_family="AluSx"
7394..7455
/rpt_family="MER5A"
7464..7627
/rpt_family="MER5B"
complement(7624..7792)
/rpt_family="MER5B"
complement(8363..8391)
/rpt_family="AT_rich"
8480..8639
/rpt_family="MER5B"
complement(9199..9286)
/rpt_family="(TA)n"
9340..9644
/rpt_family="AluB"
9880..9946
/rpt_family="MIR2"
10208..10787
/rpt_family="LMB3"
10984..11297
/rpt_family="AluSg"
11311..11470
/rpt_family="LIPB8"
complement(11471..11495)
/rpt_family="AT_rich"
complement(12384..12440)
/rpt_family="AT_rich"
13782..13921
/rpt_family="MIR"
14035..14316
/rpt_family="AluSx"
complement(15018..15046)
/rpt_family="AT_rich"
15080..15373
/rpt_family="AluB"
15586..15881
/rpt_family="AluB"
complement(18038..18071)
/rpt_family="(CAT)n"
complement(20389..20494)
/rpt_family="FAM"
21236..21534
/rpt_family="AluSx"
complement(23331..23353)
/rpt_family="GC_rich"
complement(23546..23578)
/rpt_family="GC_rich"
24565..24707
/rpt_family="MIR"
26013..26047
/rpt_family="MIR2"
complement(27717..28532)
/rpt_family="LIPB1"
complement(28553..28854)
/rpt_family="AluSg"
complement(28860..29076)
/rpt_family="L1"
complement(29365..29561)
/rpt_family="MIR"
complement(30050..30147)
/rpt_family="MIR"

```



```

repeat_region      31487..31812
                    /rpt_family="AluIo"
repeat_region      complement(33629..33732)
                    /rpt_family="MIR"
repeat_region      34434..34729
                    /rpt_family="AluSx"
repeat_region      35070..35361
                    /rpt_family="AluSx"
repeat_region      complement(36437..36461)
                    /rpt_family="AT_rich"
repeat_region      complement(39967..40261)
                    /rpt_family="AluSx"
repeat_region      complement(40741..40915)
                    /rpt_family="AluSg"
repeat_region      40963..41663
                    /rpt_family="L1PA2"
repeat_region      42042..42352
                    /rpt_family="MLT1E"
repeat_region      42467..42771
                    /rpt_family="AluSg"
repeat_region      42772..42848
                    /rpt_family="L1MB8"
repeat_region      42849..43151
                    /rpt_family="AluIb"
repeat_region      43152..43318
                    /rpt_family="L1MA10"
repeat_region      43253..43427
                    /rpt_family="L1MB8"
repeat_region      43429..43672
                    /rpt_family="L1MB8"
repeat_region      complement(44188..44481)
                    /rpt_family="AluSc"
repeat_region      44544..45246
                    /rpt_family="L1R8"
repeat_region      complement(46999..47032)
                    /rpt_family="AT_rich"
repeat_region      complement(47411..47431)
                    /rpt_family="AT_rich"
repeat_region      complement(47612..47664)
                    /rpt_family="CATRn"
repeat_region      complement(47946..48239)
                    /rpt_family="AluIo"
repeat_region      complement(49827..49872)
                    /rpt_family="AT_rich"
repeat_region      complement(50160..50186)
                    /rpt_family="AT_rich"
repeat_region      50952..51121
                    /rpt_family="MER5A"
repeat_region      complement(52140..52284)
                    /rpt_family="AluIo"
repeat_region      53772..53974
                    /rpt_family="MER1B"
repeat_region      54225..54348
                    /rpt_family="MIR"
repeat_region      complement(54533..54579)
                    /rpt_family="(TGA)n"
repeat_region      complement(54580..54736)
                    /rpt_family="AluIb"
repeat_region      complement(54755..54782)
                    /rpt_family="(CA)n"

Alignment Scores:
Pred. No.:          9.49e-07      Length:      106988
Score:              154.00        Matches:      28
Percent Similarity: 100.00%       Conservative: 0
Best Local Similarity: 100.00%     Mismatches:  0
Query Match:        100.00%        Indels:       0
DB:                  9             Gaps:         0

US-10-010-050a-2_COPY_1_28 (1-28) x ACC001226 (1-106988)
QY      1 MetAArgGjYAlaGjYAlaAlaArgGjYArgAlaSerTrrCyTrrAlaLeuAlaLeu 20
|||||

```

```

Db      23544 ATGGCGCGGGCGCGGCTCGGGACGCGCTTCTGTGCTGGCCCTGCGGCTG 23603
QY      21 LeuTrrpleuAlaValYalProGly 28
Db      23604 CTTGGCTCGCGGTGCTTCGCGG 23627

RESULT 9
AL136440
LOCUS      Homo sapiens chromosome 13 clone RP11-185124, 14 unordered pieces.
DEFINITION
ACCESSION      AL136440.2 GI:10039473
VERSION
KEYWORDS      HTG; HTGS_PHRASE1; HTGS_CANCELLED.
SOURCE
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS      Burton,J.
TITLE      Direct Submission
JOURNAL
COMMENT
On Sep 8, 2000 this sequence version replaced gi:6982057.
-----
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
Project Information
Center project name: ba185124
-----
Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; 108752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 162085 bases at least Q40
Consensus quality: 166040 bases at least Q20
Insert size: 168062; sum-of-contrigs
Insert size: 164996; 1.7% error; agarose-fp
Quality coverage: 4.08x in Q20 bases; sum-of-contrigs Quality
coverage: 4.16x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
3401 3500: contig of 3400 bp in length
*
3501 16444: contig of 12944 bp in length
*
16445 16544: gap of 100 bp
*
16545 19514: contig of 2970 bp in length
*
19515 19614: gap of 100 bp
*
19615 22241: contig of 2627 bp in length
*
22242 22341: gap of 100 bp
*
22342 25526: contig of 3185 bp in length
*
25527 25627: gap of 100 bp
*
25627 33070: contig of 7444 bp in length
*
33071 33170: gap of 100 bp
*
33171 67620: contig of 34450 bp in length
*
67621 67720: gap of 100 bp
*
67721 93413: contig of 25693 bp in length
*
93414 93513: gap of 100 bp
*
93514 114081: contig of 20568 bp in length
*
114082 114181: gap of 100 bp
*
114182 126206: contig of 12025 bp in length
*
126207 126306: gap of 100 bp
*
126307 144583: contig of 18277 bp in length
*
144584 144683: gap of 100 bp

```



55654	63088:	contig of 6545 bp	in length
63089	63188:	gap of 100 bp	
63189	65281:	contig of 2093 bp	in length
65282	65381:	gap of 100 bp	
65382	67940:	contig of 2559 bp	in length
67941	68040:	gap of 100 bp	
68041	75028:	contig of 6989 bp	in length
75030	75129:	gap of 100 bp	
75130	85667:	contig of 10438 bp	in length
85668	85667:	gap of 100 bp	
85669	89150:	contig of 3483 bp	in length
89151	89250:	gap of 100 bp	
89251	93380:	contig of 4130 bp	in length
93381	93480:	gap of 100 bp	
93481	95619:	contig of 2139 bp	in length
95620	95719:	gap of 100 bp	
95720	100599:	contig of 4873 bp	in length
100599	100692:	gap of 100 bp	
100693	103995:	contig of 3267 bp	in length
103996	140495:	gap of 100 bp	
140496	110871:	contig of 6612 bp	in length
110872	110971:	gap of 100 bp	
110972	115555:	contig of 4584 bp	in length
115556	115655:	gap of 100 bp	
115656	118593:	contig of 2338 bp	in length
118594	118693:	gap of 100 bp	
118694	123530:	contig of 4637 bp	in length
123531	123630:	gap of 100 bp	
123631	127089:	contig of 3457 bp	in length
127089	127187:	gap of 100 bp	
127188	129684:	contig of 2497 bp	in length
129685	129784:	gap of 100 bp	
129785	133132:	contig of 2248 bp	in length
133133	133232:	gap of 100 bp	
133233	147792:	contig of 11560 bp	in length
147793	143892:	gap of 100 bp	
143893	148766:	contig of 4874 bp	in length
148767	148866:	gap of 100 bp	
148867	152258:	contig of 3392 bp	in length
152259	152358:	gap of 100 bp	
152359	156922:	contig of 4564 bp	in length
156923	157022:	gap of 100 bp	
157023	160153:	contig of 3311 bp	in length
160154	160253:	gap of 100 bp	
160254	164334:	contig of 4081 bp	in length
164335	164434:	gap of 100 bp	
164435	166600:	contig of 2166 bp	in length
166601	166700:	gap of 100 bp	
166701	171896:	contig of 5196 bp	in length
171897	171996:	gap of 100 bp	
171997	184915:	contig of 12119 bp	in length
184916	185015:	gap of 100 bp	
185016	187238:	contig of 2223 bp	in length
187239	187338:	gap of 100 bp	
187339	189602:	contig of 2264 bp	in length
189603	189702:	gap of 100 bp	
189703	194463:	contig of 4761 bp	in length
194464	194563:	gap of 100 bp	
194564	197882:	contig of 3119 bp	in length
197883	197782:	gap of 100 bp	
197783	201788:	contig of 4006 bp	in length
201789	201888:	gap of 100 bp	
201889	204060:	contig of 2172 bp	in length
204061	204160:	gap of 100 bp	
204161	206237:	contig of 2077 bp	in length
206238	206337:	gap of 100 bp	
206338	210376:	contig of 4039 bp	in length
210377	210476:	gap of 100 bp	
210477	225245:	contig of 14769 bp	in length
225246	225345:	gap of 100 bp	
225346	230027:	contig of 4682 bp	in length
230028	230127:	gap of 100 bp	
230128	232610:	contig of 2483 bp	in length

FEATURES	source
232611	233710: gap of 100 bp
222711	2334942: contig of 2232 bp in length
224943	235042: gap of 100 bp
235043	240958: contig of 5116 bp in length
240959	241058: gap of 100 bp
241059	251187: contig of 10129 bp in length
	Location/Qualifiers
	1..251187

misc_feature	/organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="13" /clone="RP11-453N2" /clone_1b="RPC1-11.2" 1. 2130
misc_feature	/note="assembly_fragment:00469 fragment_chain:1" 231. .5689
misc_feature	/note="assembly_fragment:01219 fragment_chain:1" 5790. .10046
misc_feature	/note="assembly_fragment:00352 fragment_chain:1" 10147. .16624
misc_feature	/note="assembly_fragment:00547 fragment_chain:2" 12725. .20993
misc_feature	/note="assembly_fragment:00617 fragment_chain:2" 21094. .28936
misc_feature	/note="assembly_fragment:01248 fragment_chain:2" 29037. .33230
misc_feature	/note="assembly_fragment:00606 fragment_chain:3" 32331. .35153
misc_feature	/note="assembly_fragment:01016 fragment_chain:3" 35254. .37649
misc_feature	/note="assembly_fragment:01277 fragment_chain:3" 37750. .40245
misc_feature	/note="assembly_fragment:00920 fragment_chain:4" 40346. .48420
misc_feature	/note="assembly_fragment:01288 fragment_chain:4" 48521. .52169
misc_feature	/note="assembly_fragment:00615 fragment_chain:4" 52270. .56443
misc_feature	/note="assembly_fragment:00009 fragment_chain:5" 56544. .63088
misc_feature	/note="assembly_fragment:00448 fragment_chain:5" 63189. .65261
misc_feature	/note="assembly_fragment:00027

Alignment Scores:	
Pred. No.:	1,61e-05
Score:	146.00
Percent Similarity:	96.43%
Best Local Similarity:	96.43%
Query Match:	94.81%
DB:	2
	Gaps: 0
	Length: 251187
	Matches: 27
	Conservative: 0
	Mismatches: 1
	Indels: 0
	Gaps: 0

US-10-010-050A-2 COPY 1 28 (1-28) X AL359875 (1-251187)

QY	Db
1	241206
MeatargarglyalaglyalalalargglyalarglaserTropCytrTrolalaleunlaieu	ATGCGCGCGGCGCGCGCGCGCTTCGTGGGACGCGCTTCTGTGTGTGTGGGCTTGGCGCTG
20	241265



gene  
CDS  
IDLRPMLDHPISGIALHVGILYALLPFLSMALLRRRLMR"  
/complement(3222..4289)  
/gene="nodi"  
/note="Product confidence : probable"  
/function="cell processes; nodulation"  
/note="Product confidence : probable"  
Gene name confidence : putative"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="PROBABLE NODI MEMBRANE TRANSPORT PROTEIN"  
/protein\_id="CAD31532.1"  
/db\_xref="GI:20803955"  
/db\_xref="GOA:P23703"  
/db\_xref="GOA:Q8KJ16"  
/db\_xref="SWISS-PROT:P23703"  
/translation="MISGRVLSQDQHGVYLRKLGSGEDLRRETPAIEREFHQISA  
KSSVPSLSTAVDVFAGVTSYGNKIYVDELSFVSGECFGLPGAGKSTIAEML  
LGMTCPPAGTTTVGVPPAPARARLRGIGVPOFMDLDOEFTVRENLVGRYFGMS  
TRGEAVTPSLIEPARLERKADARVSLSGMRCCLMARLINDPOLYMDPTTGL  
DPARHLIWERLRLARLKGKTLITTPMEBAELCDRLVLEKGRNIAAGCPQALID  
EHIGQYMEIYVGDPSHRLSLVXPHSQRISVSEETLYCVAPDPQVRYTOLQGRARL  
LIRPANEDVFLRLTGREMEB"  
/complement(4294..5568)  
/gene="nodi"  
/complement(4294..5568)  
/gene="nodi"  
/function="cell processes; nodulation"  
/note="Product confidence : probable"  
Gene name confidence : putative"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="PROBABLE NODULATION PROTEIN C,  
N-ACETYLGLUCOSAMINYLTRANSFERASE"  
/protein\_id="CAD31533.1"  
/db\_xref="GI:20803956"  
/db\_xref="GOA:Q8KJ15"  
/db\_xref="SPRMBL:Q8KJ15"  
/translation="MTLPATASVAICGVALLSVYKTAQVFTLPNNVPTSDPPG  
GEPWSPVILIPCNVEARPTLSDCLASISQGVAGKLQVYVVDGSSARDELVGHEB  
YAGDPREVPALPKVNGRKAQIAIRSCDLYLVNDSDYILAPDVTTRALMDQO  
AVGAMQOLASNSSETMLRLIDMEVYLACNEERAAARGAVMCCGPCAMTRISA  
LVSLIDQYETORFGRKSPDPRGDRHLLTIMKAGRRYVEAVAAVTPNSMPYLR  
QOLRVARSTFRDTLAFOLRLGNLYTLIDYIGNTGILLSLISLAGLADLYTGTG  
PWTACLMIAAMTYVRCVAAFRAQLRGLFSHLTLNIFILPILKAYALCTLSN3DM  
LSRSSAAVTDYGTSSKPNLVGSDAAYSEQO"  
/complement(5624..6220)  
/gene="noda"  
/complement(5624..6220)  
/gene="noda"  
/function="cell processes; nodulation"  
/note="Product confidence : probable"  
Gene name confidence : putative"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="PROBABLE NODA N-ACYLTRANSFERASE PROTEIN"  
/protein\_id="CAD31534.1"  
/db\_xref="GI:20803957"  
/db\_xref="GOA:Q8KJ14"  
/db\_xref="SPRMBL:Q8KJ14"  
/translation="KRPDYQRLCMENELQDPDHLISFPFKTYEPGAGNKGQFEG  
GSSWAGARPELRAIGDVHGVAAHMGVLRRTYKVGDDYDLVAVELGYGRPDLBGLI  
AHGISYMPVLQOLGVPAFGVTRQALRNHVSFCKGILANTLTGVRASTLDPVYSD  
LPSITRVDYDVLNVLPIGRSMSEWPAITLIDRNGPEL"  
/complement(6579..6833)  
/gene="msl130"  
/complement(6579..6833)  
/gene="msl130"  
/function="miscellaneous; unknown"  
/note="Product confidence : hypothetical"

gene  
CDS  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="HYPOTHETICAL PROTEIN"  
/protein\_id="CAD31535.1"  
/db\_xref="GI:20803958"  
/db\_xref="SPRMBL:Q8KJ17"  
/translation="MALAQWLAGGQKQVRSSIALNRRLVNAFECRSTGLSVKAP  
GREGARPDGATSNELFEVYADWNTQLREHREVRKAGP"  
/complement(7450..8058)  
/gene="nods"  
/complement(7450..8058)  
/gene="nods"  
/function="cell processes; nodulation"  
/note="Product confidence : probable"  
Gene name confidence : putative"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="PROBABLE NODULATION PROTEIN S,  
METHYLTRANSFERASE"  
/protein\_id="CAD31536.1"  
/db\_xref="GI:20803959"  
/db\_xref="GOA:Q8KJ13"  
/db\_xref="SPRMBL:Q8KJ13"  
/translation="MTHKIDHELRNEMADDEWRLDGNEFERERTOLRLSLAQR  
VANALRVGCAAGAFETRLAICOLTVIVVESEALIRARLMEKTEPHIISIDVOOF  
SSDELFDLYVAEVLVYVGDLDIIRAVRNVLISLARGCOLVGSARDANCORWGHAA  
GATTYAMNESITETIECTECRGASINDECLALFRPLIAS"  
/complement(8659..9840)  
/gene="msl132"  
/complement(8659..9840)  
/gene="msl132"  
/function="small molecule metabolism"  
/note="Product confidence : putative"  
Gene name confidence : hypothetical"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=1  
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"  
/protein\_id="CAD31537.1"  
/db\_xref="GI:20803960"

Alignment Scores:  
Pred. No.: 1,85e+03 Length: 152050  
Score: 73.00 Matches: 12  
Percent Similarity: 70.00% Conservative: 2  
Best Local Similarity: 60.00% Mismatches: 6  
Query Match: 47.40% Indels: 0  
DB: 1 Gaps: 0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x ML0672113 (1-152050)

QY 1 Metltargaggglyalalalargylasertppcystipalatalaenu 20  
DB 130027 ATGCAGACGAGCGGAGAGGTGTGCGGTGAGAGCGCATGTGTGAGCTTACGCCTC 129968  
||||:|||||  
|||||

RESULT 13  
HSBA19D2/C 166913 bp DNA linear PRI 10-MAR-2001  
LOCUS Human DNA sequence from clone Rpl1-19D2 on chromosome 20. Contains  
DEFINITION a putative novel gene, ESTs, STSs and GSSs, complete sequence.  
ACCESSION AL080248  
VERSION AL080248.7 GI:5725257  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 166913)  
AUTHORS McMurray A.  
TITLE Direct Submission



```

misc_feature      /replace="atc"
                  20823..21222
                  /note="match: STS: Em:G18104"
misc_feature      21998..22438
                  /note="match: STS: Em:AQ476128"
repeat_region     23244..23380
                  /note="MIR repeat: matches 49..181 of consensus"
variation          23350..23352
                  /note="clone RPS-988617
                  cta in this entry
                  substitution"
                  /replace="caa"
repeat_region     23559..23967
                  /note="MIR1B repeat: matches 1..390 of consensus"
variation          24007..24009
                  /note="clone RPS-988617
                  agt in this entry
                  substitution"
                  /replace="aat"
misc_feature      join(24433..24663,25113..25163)
                  /note="match: GSS: Em:AL023725"
misc_feature      complement(join(24573..24708,82902..82963))
                  /note="match: GSS: Em:AQ315747"
repeat_region     24729..25068
                  /note="MER7A repeat: matches 2..346 of consensus"
misc_feature      complement(join(25074..25207,78259..78297))
                  /note="match: STS: Em:Z523394"
misc_feature      join(25143..25339,82904..82972)
                  /note="match: STS: Em:G07821"
repeat_region     25388..25620
                  /note="MER44A repeat: matches 1..259 of consensus"
variation          25571..25573
                  /note="clone RPS-988617
                  cgt in this entry
                  substitution"
                  /replace="cat"
repeat_region     25827..26055
                  /note="MER44A repeat: matches 503..333 of consensus"
variation          25935..25937
                  /note="clone RPS-988617
                  taa in this entry

Alignment Scores:
Pred. No.:      2,61e+03      Length:      166913
Score:          72.00         Matches:      12
Percent Similarity: 60.87%      Conservative: 2
Best Local Similarity: 52.17%      Mismatches: 9
Query Match:    46.75%         Indels:      0
DB:             9             Gaps:        0

US-10-010-050A-2_COPY_1_28 (1-28) x HSBAL1D2 (1-166913)
Qy      4 G1YALAG1YALAA1ARG1YARG1ASERTPYCTTPALALEU1ALEU1TTPLEU 23
      Db 156648 GGGTCAGCTTCGCAAGAGTGGTCCATGTGTGGACACGCTCCACTTCAGTGGCTT 156589
      Qy      24 A1A1A1A1 26
      Db 156588 TGCAGGGTA 156580

RESULT 14
AC079116/c      AC079116      147856 bp      DNA      linear      HTG 23-SEP-2000
LOCUS           AC079116      SEQUENCE, 9 unordered pieces.
DEFINITION      AC079116.2 GI:10280934
ACCESSION       AC079116.2 GI:10280934
VERSION         AC079116.2 GI:10280934
KEYWORDS        HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE          Homo sapiens (human)
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 147856)

```

AUTHORS  
Waterston,R.H.  
TITLE  
The sequence of Homo sapiens clone  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 147856)  
AUTHORS  
Waterston,R.H.  
TITLE  
Direct Submission  
Submitted (18-AUG-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT  
On Sep 23, 2000 this sequence version replaced gi:3845173.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0103D09
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid, 0%
Chemistry: Dye-primer ET, 100% of reads
Chemistry: Dye-terminator Big Dye, 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 144423 bases at least Q40
Consensus quality: 145315 bases at least Q30
Insert size: 162000; agarose-fp
Insert size: 146912; sum-of-contigs
Quality coverage: 5.04 in Q20 bases; agarose-fp
Quality coverage: 5.59 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1731: contig of 1731 bp in length
* 1732 1831: gap of unknown length
* 1832 2887: contig of 1056 bp in length
* 2888 2987: gap of unknown length
* 2988 3131: contig of 144 bp in length
* 3132 3231: gap of unknown length
* 3232 10519: contig of 7288 bp in length
* 10520 27623: contig of 17004 bp in length
* 27624 27723: gap of unknown length
* 27724 51531: contig of 23808 bp in length
* 51532 51631: gap of unknown length
* 51632 78470: contig of 26839 bp in length
* 78471 78570: gap of unknown length
* 78571 112402: contig of 33832 bp in length
* 112403 112502: gap of unknown length
* 112503 147856: contig of 35354 bp in length.
Location/Qualifiers
1..147856
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="YXK"
/clone="RP11-103D9"
1..1731
/note="assembly_name:Contig45"
1832..2887
/note="assembly_name:Contig46"
clone_end:SP6
vector_side:left"
2988..3131
/note="assembly_name:Contig44"
3232..10519
/note="assembly_name:Contig47"

```





```

repeat_region /rpt_family="MIR"
9105. .9309
/rpt_family="L2"
repeat_region /rpt_family="L2"
9315. .9447
/rpt_family="L2"
repeat_region /rpt_family="MIR"
9848. .9894
/rpt_family="L1"
repeat_region /rpt_family="L1"
10033. .10522
/rpt_family="L1"
repeat_region /rpt_family="ERVU"
10523. .10935
/rpt_family="ERVU"
repeat_region /rpt_family="ERVU"
11051. .11107
/rpt_family="ERVU"
repeat_region /rpt_family="L1"
11109. .14499
/rpt_family="L1"
repeat_region /rpt_family="ERVU"
14500. .14593
/rpt_family="ERVU"
repeat_region /rpt_family="ERVU"
14594. .14945
/rpt_family="ERVU"
repeat_region /rpt_family="L1"
14946. .16084
/rpt_family="L1"
repeat_region /rpt_family="MaLR"
16086. .16442
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
16443. .16815
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
16816. .17148
/rpt_family="MaLR"
repeat_region /rpt_family="L1"
17151. .17355
/rpt_family="L1"
repeat_region /rpt_family="L1"
17498. .17933
/rpt_family="L1"
repeat_region /rpt_family="L1"
17993. .18179
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
20280. .20344
/rpt_family="MIR"
repeat_region /rpt_family="ERVU"
20349. .20724
/rpt_family="ERVU"
repeat_region /rpt_family="ERVU"
20723. .20765
/rpt_family="Alu"
repeat_region /rpt_family="Alu"
20786. .20890
/rpt_family="ERVU"
repeat_region /rpt_family="MIR"
20920. .21040
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
21846. .21922
/rpt_family="L2"
repeat_region /rpt_family="MIR"
22488. .22592
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
25690. .25818
/rpt_family="L2"
repeat_region /rpt_family="L2"
26126. .26425
/rpt_family="MaLR"
repeat_region /rpt_family="MaLR"
26689. .26911
/rpt_family="L2"
misc_feature /note="similar to EST A1316470 (NID:G4031737) u560f10.y1"
30862. .31090
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
31559. .31806
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
31862. .32102
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
33621. .33835
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
34010. .34116
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
34221. .34300
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
35380. .35498
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
35537. .35572
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
38009. .38147
/rpt_family="MIR"
repeat_region /rpt_family="MIR"
38409. .38767
/rpt_family="MIR"

```

```

repeat_region 38946. .39246
/rpt_family="Alu"
repeat_region 39528. .39830
/rpt_family="Alu"
repeat_region 39982. .40376
/rpt_family="L2"
repeat_region 40384. .40514
/rpt_family="Alu"
repeat_region 41336. .41487
/rpt_family="MIR"
repeat_region 42204. .42336
/rpt_family="MIR"
repeat_region 42527. .42744
/rpt_family="MIR"

```

## Alignment Scores:

Pred. No.:	Length:	Matches:
Score: 71.00	152996	13
Percent Similarity: 60.87%	Conservative: 1	
Best Local Similarity: 56.52%	Mismatches: 9	
Query Match: 46.10%	Indels: 0	
DB: 9	Gaps: 0	

US-10-010-050a-2\_copy\_1\_28 (1-28) x AC008177 (1-152996)

```

QY      4 G1YAlAG1YAlA1AArgG1YArgAlaseRTTCYGTTPAlaleuAlaleuLeuTTPleu 23
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      105444 GGTCATGCTGATGCAGAGGTGGGTTTCACTGCTGTGGGAGCTCTACTCTATGACTT 105385
QY      24 Alavalval 26
         ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      105384 TGCAGGCTA 105376

```

Search completed: April 25, 2004, 03:24:31  
 Job time : 620.278 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

1 protein - nucleic search, using frame\_plus\_p2n model

on: April 24, 2004, 22:59:32 ; Search time 567.083 Seconds

(without alignments)  
2292.945 Million cell updates/sec

US-10-010-050a-2\_COPY\_1\_30

File: 169  
Sequence: 1 MRRGAGAGARGASWCWALALIMLAVVPGWS 30

oring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

arched: 3470272 seqs, 21671516395 residues

otal number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

set-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODE=frame+ p2n.model -DEV=xlh  
3=/cgen2.1/USPTO.spool/US10010050/runat.22042004.113204.27549/app.query.fasta.1.1372  
DB=GenEmbl -QPWT=fastp -SUFFIX=rge -MINMATCH=0.1 -IOOPT=0 -IOEXT=45  
NITS=bites -START=1 -MATRIX=biosum62 -TRANS=human4.cdi -LIST=45  
DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
JREF=US10010050 @CGN 1.1 7433 @runat.22042004.113204.27549 -NCPU=6 -ICPU=3  
NO MAP -LARGEOUTPUT -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=5  
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

atabase:

GenEmbl:1:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rdt:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrc:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	169	100.0	697	6	BD125187	BD125187 Primer fo
2	169	100.0	697	6	BD126453	BD126453 Primer fo
3	169	100.0	1486	6	BD073402	BD073402 Secreted
4	169	100.0	1751	6	BD136338	BD136338 95 human
5	169	100.0	2120	6	BD127858	BD127858 Primer fo
6	169	100.0	2120	9	AK075109	AK075109 Homo sapi
7	169	100.0	4080	9	AF068227	AF068227 Homo sapi
8	169	100.0	106988	9	AC001226	AC001226 Genomic s
9	169	100.0	169362	2	AL136440	AL136440 Homo sapi
10	169	100.0	169362	2	AL1359875	AL1359875 Homo sapi
11	117	69.2	1038	6	BD073413	BD073413 Secreted
12	78.5	46.4	686	11	PM2H2B	PM2H2B
13	77.5	45.9	28835	9	HS444G9	HS444G9
14	77.5	45.9	256073	9	AE006464	AE006464 Homo sapi
15	76	45.0	105779	9	AL137007	AL137007 Human DNA
16	75	44.4	73807	9	AL606530	AL606530 Human DNA
17	74	43.8	5214	9	AL627434	AL627434 Human DNA
18	74	43.8	290625	2	AL442068	AL442068 Homo sapi
19	73	43.2	118955	9	AC068274	AC068274 Homo sapi
20	73	43.2	152050	1	ML0672113	ML0672113 Mesothiz
21	73	43.2	191072	2	AC115775	AC115775 Mus muscu
22	73	43.2	229553	2	AC115773	AC115773 Mus muscu
23	72.5	42.9	186359	10	AL596209	AL596209 Mouse DNA
24	72.5	42.9	188202	2	AC084044	AC084044 Mus muscu
25	72	42.6	100072	9	AL137144	AL137144 Human DNA
26	72	42.6	166913	9	HSBA19D2	HSBA19D2 Human DNA
27	71.5	42.3	249773	2	AC103064	AC103064 Rattus no
28	71.5	42.3	321034	2	AC133998	AC133998 Rattus no
29	71	42.0	13897	1	ABCL1SDA	ABCL1SDA
30	71	42.0	56050	9	AC097719	AC097719 Homo sapi
31	71	42.0	145969	9	AC025028	AC025028 Homo sapi
32	71	42.0	147856	2	AC079116	AC079116 Homo sapi
33	71	42.0	152996	9	AC008177	AC008177 Homo sapi
34	71	42.0	153740	2	AC027196	AC027196 Homo sapi
35	71	42.0	168467	9	AL450346	AL450346 Human DNA
36	71	42.0	174124	2	AC021386	AC021386 Homo sapi
37	71	42.0	184406	9	AC009476	AC009476 Homo sapi
38	71	42.0	201535	9	AC007248	AC007248 Homo sapi
39	71	42.0	344805	1	BX640434	BX640434 Bortdell
40	71	42.0	348074	1	BX640449	BX640449 Bortdell
41	71	42.0	348134	1	BX640420	BX640420 Bortdell
42	70.5	41.7	759	8	AK102002	AK102002 Oryza sat
43	70.5	41.7	1028	8	AK059403	AK059403 Oryza sat
44	70.5	41.7	56975	8	AC087615	AC087615 Homo sapi
45	70.5	41.7	153732	2	AC132883	AC132883 Mus muscu

RESULT 1

## ALIGNMENTS

	BID125187	697 bp	DNA	linear	PAT 18-SEP-2002
CDCS .					
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BID125187				
VERSION	BID125187.1	GI:22220132			
TIMWORDS	JP 2002017375-A/618.				
SOURCE	Homo sapiens (human)				
ORGANISM					
EREFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.				
TITLE	Primer for synthesizing full-length CDNA and use thereof				
JOURNAL	Patent: JP 2002017375-A 618 22-JAN-2002; HELIX RESEARCH INSTITUTE				
COMMENT	OS Homo sapiens (human) PN JP 2002017375-A/618 PD 22-JAN-2002 PE 07-JUL-2000 JP 2000253172 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO PI YORI KAWAI, AI WAKAMATSU TOMOKASU SUGIYAMA, KEICHI NAGAI, PI SHINICHI KOJIMA, PI TETSUO OTSUKI, HISASHI KOGA PC C12N15/09,C07KL4/47,C07KL6/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC C12N15/09,C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC '10' Primer for synthesizing full-length cDNA and use thereof FH Key  location/Qualifiers FT source 1..697 /organism='Homo sapiens (human)'. FT location/Qualifiers 1..697 /organism='Homo sapiens' /mol_type='genomic DNA' /db_xref='taxon:9606'				
FEATURES					
source					
ORIGIN					
Alignment Scores:					
Pred. No.:	2.69e-10	length:	697		
Score:	163.00	Matches:	30		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	100.00%	Indels:	0		
DB:	6	Gaps:	0		
US-10-010-050A-2_COPY_1_30 (1-30) x BID125187 (1-697)					
QY	1 MetaArgArGgAlGaLgaLaalaaRGGLVaRgaLasErTrPyCTTPPaLeuaLaLu 2)				
Dd	52 ATGGCGGC GG GCG CG GC GG GCG GTT CG GG GA CG CCTTC TG TC TTGG CTGG GCCCTTG A 111				
QY	21 LeuTrPLEuaLaValvaLPProGLTYrSer 30				
Dd	112 CTTTGCCGTGCGGGTGGTTCGCGGCCGCTGCTGC 141				
RESULT 2					
Locus	BID126453	697 bp	DNA	linear	PAT 18-SEP-2002
DEFINITION	Primer for synthesizing full-length cDNA and use thereof.				
ACCESSION	BID126453				
VERSION	BID126453.1	GI:23221398			
KEYWORDS	JP 2002017375-A/1884.				
SOURCE	Homo sapiens (human)				
ORGANISM					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
	(bases 1 to 697)				
	Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y., Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and Koga,H.				
REFERENCE	PAT 18-SEP-2002				
AUTHORS	OTA,T., NISHIKAWA,T., ISOgai,T., HayaSHI,K., IShII,S., Kawai,Y., WakamatsU,A., Sugiyama,T., NaGal,K., KoJIma,S., OtSuKi,T. and				

TITLE	JOURNAL	COMMENT
Koga, H.		
Primer for synthesizing full-length cDNA and use thereof		
Patent: JP 2002017375-A 1884 22-JAN-2002;		
HELIX RESEARCH INSTITUTE		
OS Homo sapiens (human)		
PN JP 2002017375-A/1884		
PD 22-JAN-2002		
PF 07-JUL-2000 JP 2000253172		
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO		
PI ISHII,		
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI		
SHINICHI KOGIMA,		
PI TETSUO IOTSUKI, HISASHI KOGA		
PC C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/		
10, PC C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC		
Primer for synthesizing full-length cDNA and use thereof FH Key		
Location/Qualifiers		
FT source 1..697		
Location/Qualifiers		
1..697		
Location/Qualifiers		
1..697		
/organism="Homo sapiens"		
/mol_type="genomic DNA"		
/db_xref="taxon:9606"		

[illegible]

		CC n equals a,t,g, or c			
		CC n equals a,t,g, or c			
FH	Key	Location/Qualifiers			
FT	source	1..1751			
		/organism='Homo sapiens (human)'			
FEATURES					
	source	1..1751			
		/organism='Homo sapiens"			
		/mol_type='genomic DNA"			
		/db_xref='taxon:9606"			
ORIGIN					
	Alignment Scores:				
	Pred. No.:	6.29e-10	Length:	1751	
	Score:	169.00	Matches:	30	
	Percent Similarity:	100.00%	Conservative:	0	
	Best local Similarity:	100.00*	Mismatches:	0	
	Query Match:	100.00*	Indels:	0	
	DB:	6	Gaps:	0	
	US-10-010-050A-2_COPY_1_30 (1-30) x BD136338 (1-1751)				
OY		1 MetArgArgGlyIValIacIyAlaAlaArGlyARgAlaserIPcySTrPalaleuAlau 20			
Dd		49 ATCGGGGGGGGGGGGGGGGGCTTTCGGAGACGCCCTTGTTGTGCTGGCCTGGCGCTG 108			
OY		21 LeuTrIPleuaIAVaIVaIProGIYTTPSer 30			
Dd		109 CTTTGCTCGGGGTGGTTCGGGCTGGATCC 138			
RESULT 5					
BD127858					
LOCUS		BD127858	2120 bp	DNA	linear PAT 18-SEP-2002
DEFINITION					
ACCESSION		BD127858			
VERSION		BD127858.1 GI:23222803			
KEYWORDS		JP 2002017375-A/3289.			
SOURCE		Homo sapiens (human)			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
		1 (bases 1 to 2120)			
		Oca,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,			
		Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and			
		Koga,H.			
TITLE		Primer for synthesizing full-length cDNA and use thereof			
JOURNAL		Patent: JP 2002017375-A 3289 22-JAN-2002;			
COMMENT		HEALTH RESEARCH INSTITUTE			
		OS Homo sapiens (human)			
		PN JP 2002017375-A/3289			
		PD 22-JAN-2002			
		PF 07-JUL-2000 JP 2000253172			
		PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO			
		PI ISHII,			
		PI YURI IKAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI MAGAI, PI			
		SHINICHI KOJIMA,			
		PI TETSUJI OTSUKI,HISASHI KOGA			
		PC			
		C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC			
		'10,			
		PC C12P21/02,C12Q1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC			
		Primer for synthesizing full-length cDNA and use thereof FH Key			
		Location/Qualifiers			
		(16)..(1089).			
FEATURES					
	source	1..2120			
		/organism='Homo sapiens'"			
		/mol_type='genomic DNA"			
		/db_xref='taxon:9606"			
ORIGIN					
	Alignment Scores:				
	Pred. No.:	7.5e-10	Length:	2120	



21 LeuTripleAlaValValProGlyTyrSer 30  
 |||||  
 1536 CTTGGCTCGCGGTGTTCGGGCTGGTCC 1565

SUIT 8  
 001226

CDS  
 AC001226 106988 bp DNA linear PRI 29-MAY-1997  
 Genomic sequence from Human 13, complete sequence.

FINITION  
 AC001226

CESSION  
 AC001226.1 GI:2133862

RSION  
 HTG.

YWORDS  
 Homo sapiens (human)

ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 Hawkins, K.H., Reeve, M.P., Christoffersen, A., Birren, B.W.,  
 1 (bases 1 to 106988)

AUTHORS  
 Fasmann, K.H. and Lander, E.S.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Genomic sequence from Human 13  
 Unpublished  
 2 (bases 1 to 106988)  
 Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,  
 Fasmann, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,  
 Barina, N., Brown, K., Cooke, P., Daly, M.J., Forrest, C., Frapp, W.J.,  
 Gage, D., Geradigery, K., Hagos, B., Jacotot, L., Lane, M., Mackenzie, J.,  
 Margulis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A.,  
 Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J.,  
 Stilwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and  
 Zody, M.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Direct Submission  
 Submitted (10-APR-1997) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 106988)  
 Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,  
 Fasmann, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,  
 Barina, N., Chang, A., Cooke, P., Daly, M.J., Forrest, C., Frapp, W.J.,  
 Gage, D., Geradigery, K., Hagos, B., Jacotot, L., Lane, M., Mackenzie, J.,  
 Margulis, N., McDermott, J., Moloney, N., Morrow, J., Nachman, A.,  
 Naylor, J., O'Connor, T., Peterson, K., Rollins, G., Spencer, J.,  
 Stilwell, J., Stone, C., Strickland, C., Sydney, K., Wilmer, F. and  
 Zody, M.

TITLE  
 JOURNAL  
 COMMENT  
 Direct Submission  
 Submitted (29-MAY-1997) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On May 29, 1997 this sequence version replaced gi:1932720.  
 The Staden databases, finishing information, and all  
 chromatographic files used in the assembly of this clone are  
 available from our anonymous ftp site.

FEATURES  
 source  
 All repeats were identified using RepeatMasker: Smit, A.F.A. &  
 Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html.  
 location/Qualifiers  
 1..106988  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /clone="76N15"  
 /clone\_lib="unknown"  
 /complement(113..200)  
 /rpt\_family="THE1B"  
 /rpt\_family="L1PA3"  
 /rpt\_family="L1PA3"  
 /complement(674..694)  
 /rpt\_family="AT\_rich"  
 /complement(835..981)  
 /rpt\_family="MIR2"  
 /rpt\_family="L1PA3"  
 /complement(1632..1920)  
 /rpt\_family="AluJo"  
 /rpt\_family="AluJo"  
 /complement(1970..2271)  
 /rpt\_family="AluSp"  
 /rpt\_family="AluSp"  
 /complement(2779..2946)  
 /rpt\_family="MIR"

repeat\_region /rpt\_family="FAM"  
 3552..3828  
 /rpt\_family="AluDb"  
 repeat\_region 4361..4640  
 /rpt\_family="AluSg"  
 repeat\_region 5253..5385  
 /rpt\_family="MLT1F"  
 repeat\_region 5914..6106  
 /rpt\_family="MLT1F"  
 repeat\_region complement(6341..6435)  
 /rpt\_family="MER34"  
 repeat\_region complement(6838..7140)  
 /rpt\_family="AluSx"  
 repeat\_region 7394..7455  
 /rpt\_family="MERSA"  
 repeat\_region 7464..7627  
 /rpt\_family="MERSB"  
 repeat\_region complement(7624..7792)  
 /rpt\_family="MERSB"  
 repeat\_region complement(8363..8391)  
 /rpt\_family="AT\_rich"  
 repeat\_region 8480..8639  
 /rpt\_family="MERSB"  
 repeat\_region complement(9199..9286)  
 /rpt\_family="MERSB"  
 repeat\_region 9340..9644  
 /rpt\_family="AluJo"  
 repeat\_region 9880..9946  
 /rpt\_family="MIR2"  
 repeat\_region 10208..10787  
 /rpt\_family="L1MB3"  
 repeat\_region 10984..11297  
 /rpt\_family="AluSg"  
 repeat\_region 11311..11470  
 /rpt\_family="L1PA8"  
 repeat\_region complement(11471..11495)  
 /rpt\_family="AT\_rich"  
 repeat\_region complement(12394..12440)  
 /rpt\_family="AT\_rich"  
 repeat\_region 13782..13921  
 /rpt\_family="MIR"  
 repeat\_region 14035..14316  
 /rpt\_family="AluSx"  
 repeat\_region complement(15018..15046)  
 /rpt\_family="AT\_rich"  
 repeat\_region 15080..15373  
 /rpt\_family="AluDb"  
 repeat\_region 15588..15881  
 /rpt\_family="AluJo"  
 repeat\_region complement(18038..18071)  
 /rpt\_family="CATn"  
 repeat\_region complement(20389..20494)  
 /rpt\_family="FAM"  
 repeat\_region 21238..21534  
 /rpt\_family="AluSx"  
 repeat\_region complement(23331..23353)  
 /rpt\_family="GC\_rich"  
 repeat\_region complement(23546..23578)  
 /rpt\_family="GC\_rich"  
 repeat\_region 24565..24707  
 /rpt\_family="MIR"  
 repeat\_region 26013..26047  
 /rpt\_family="MIR2"  
 repeat\_region complement(27717..28532)  
 /rpt\_family="L1PB1"  
 repeat\_region complement(28553..28854)  
 /rpt\_family="AluSg"  
 repeat\_region complement(28860..29076)  
 /rpt\_family="L1A"  
 repeat\_region complement(29365..29561)  
 /rpt\_family="MIR"  
 repeat\_region complement(30050..30147)  
 /rpt\_family="MIR"







```
* 56544 63088: contig of 6545 bp in length
* 63089 63188: gap of 100 bp
* 63189 65281: contig of 2093 bp in length
* 65282 65381: gap of 100 bp
* 65382 67940: contig of 2559 bp in length
* 67941 68040: gap of 100 bp
* 68041 75029: contig of 6988 bp in length
* 75030 75129: gap of 100 bp
* 75130 85567: contig of 10438 bp in length
* 85568 85667: gap of 100 bp
* 85668 89150: contig of 3483 bp in length
* 89151 89250: gap of 100 bp
* 89251 93380: contig of 4130 bp in length
* 93381 93480: gap of 100 bp
* 93481 95619: contig of 2139 bp in length
* 95620 95719: gap of 100 bp
* 95720 100592: contig of 4873 bp in length
* 100593 100692: gap of 100 bp
* 100693 103959: contig of 3267 bp in length
* 103960 104059: gap of 100 bp
* 104060 110871: contig of 6812 bp in length
* 110872 110971: gap of 100 bp
* 110972 115555: contig of 4584 bp in length
* 115556 115655: gap of 100 bp
* 115656 118593: contig of 2938 bp in length
* 118594 118693: gap of 100 bp
* 118694 123530: contig of 4837 bp in length
* 123531 123630: gap of 100 bp
* 123631 127087: contig of 3457 bp in length
* 127088 127187: gap of 100 bp
* 127188 129684: contig of 2497 bp in length
* 129685 129784: gap of 100 bp
* 129785 132132: contig of 2348 bp in length
* 132133 132232: gap of 100 bp
* 132233 143792: contig of 11560 bp in length
* 143793 143892: gap of 100 bp
* 143893 148766: contig of 4874 bp in length
* 148767 148866: gap of 100 bp
* 148867 152258: contig of 3392 bp in length
* 152259 152358: gap of 100 bp
* 152359 156922: contig of 4564 bp in length
* 156923 157022: gap of 100 bp
* 157023 160153: contig of 3131 bp in length
* 160154 160253: gap of 100 bp
* 160254 164334: contig of 4081 bp in length
* 164335 164434: gap of 100 bp
* 164435 166700: contig of 2166 bp in length
* 166701 166800: gap of 100 bp
* 168001 171896: contig of 5196 bp in length
* 171897 171996: gap of 100 bp
* 171997 184915: contig of 12919 bp in length
* 184916 185015: gap of 100 bp
* 185016 187238: contig of 2223 bp in length
* 187239 187338: gap of 100 bp
* 187339 189602: contig of 2264 bp in length
* 189603 189702: gap of 100 bp
* 189703 194463: contig of 4761 bp in length
* 194464 194563: gap of 100 bp
* 194564 197682: contig of 3119 bp in length
* 197683 197782: gap of 100 bp
* 197783 201788: contig of 4006 bp in length
* 201789 201888: gap of 100 bp
* 201889 204060: contig of 2172 bp in length
* 204061 204160: gap of 100 bp
* 204161 206237: contig of 2077 bp in length
* 206238 206337: gap of 100 bp
* 206338 210376: contig of 4039 bp in length
* 210377 210476: gap of 100 bp
* 210477 225245: contig of 14769 bp in length
* 225246 225345: gap of 100 bp
* 225346 23027: contig of 4682 bp in length
* 23028 23027: gap of 100 bp
* 230128 232610: contig of 2483 bp in length
```

## FEATURES

source

```
* 232611 232710: gap of 100 bp
* 232711 234942: contig of 2232 bp in length
* 234943 235042: gap of 100 bp
* 235043 240958: contig of 5916 bp in length
* 240959 241058: gap of 100 bp
* 241059 251187: contig of 10129 bp in length.
Location/Qualifiers
1. .251187
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone_11b="RPCT-11.2"
1. .2130
/note="assembly fragment:00469
fragment chain:1"
misc_feature
2231. .5689
/note="assembly fragment:01219
fragment chain:1"
misc_feature
5790. .10046
/note="assembly fragment:00352
fragment chain:1"
10147. .12624
/note="assembly fragment:00547
fragment chain:2"
misc_feature
12725. .20993
/note="assembly fragment:00617
fragment chain:2"
21094. .78936
/note="assembly fragment:01248
fragment chain:2"
misc_feature
29037. .32230
/note="assembly fragment:00606
fragment chain:3"
32331. .35153
/note="assembly fragment:01016
fragment chain:3"
35254. .37649
/note="assembly fragment:01277
fragment chain:3"
37750. .40245
/note="assembly fragment:00920
fragment chain:4"
40346. .48420
/note="assembly fragment:01288
fragment chain:4"
48521. .52169
/note="assembly fragment:00615
fragment chain:4"
misc_feature
52270. .56443
/note="assembly fragment:00009
fragment chain:5"
56544. .63088
/note="assembly fragment:00448
fragment chain:5"
63189. .65281
/note="assembly fragment:00027
fragment chain:5"
misc_feature
```

## Alignment Scores:

```
Pred. No.: 4.93e-07 Length: 251187
Score: 161.00 Matches: 29
Percent Similarity: 96.67% Conservative: 0
Best Local Similarity: 96.67% Mismatches: 1
Query Match: 95.27% Indels: 0
DB: 2 Gaps: 0
```

US-10-010-050A-2\_COPY\_1\_30 (1-30) x AL359875 (1-251187)

Oy 1 MetARGAGGlyAlaGlyAlaArgGlyArgAlaSerTPCySTPAlaLeuAlaLeu 20

Db 241206 ATGGCGGGGGGGGGGGGGGGCTCGGAGAACGCGCTTCCTGCTGCGGCGCTG 241265



assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep/la16c-444g9](http://www.sanger.ac.uk/Projects/C_elegans/wormpep/la16c-444g9) is part of a clone contig from the tip of the short arm of chromosome 16 spanning 2Mb of p13.3 (Higgs D.R., Flint J., Daniels R., MRC Molecular Haematology Unit, Institute of Molecular Medicine, Oxford (unpublished)), and is from the Los Alamos, flow sorted human chromosome 16 libraries constructed by Norman Doggett (unpublished). VECTOR: sCos-1.

## FEATURES

source

1. 28835  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="LA16c-444G9"  
/clone\_id="LA16"

## ORIGIN

Alignment Scores:

Ref. No.:	176	Length:	28835
Percent Similarity:	77.50	Matches:	15
Percent Local Similarity:	59.26%	Conservative:	1
Percent Match:	55.56%	Mismatches:	10
	45.86%	Indels:	1
		Gaps:	1

US-10-010-050a-2\_COPY\_1\_30 (1-30) x HS444G9 (1-28835)

Y 4 GlyAlaGlyAlaAlaAlaG--GlyArgAlaSerTrpCysTrpAlaLeuAlaLeuTrp 22

Jb 27519 GGGGCTGGGCTGGCGCCGAGGCGGGCGGTGGAGTGGGCTGTAGTCCGGTGG 27578

Y 23 LeuAlaValValProGlyTrp 29

Jb 27579 ACGGGATGCAAGCCAGCTGG 27599

## RESULT 14

JOCUS AE006464 256073 bp DNA linear PRI 27-AUG-2002

DEFINITION Homo sapiens 16p13.3 sequence section 3 of 8.

ACCESSION AE006464 AE005175

VERSION AE006464.1 GI:14336700

## KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 256073)  
Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,  
Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and  
Higgs,D.R.

TITLE Sequence, structure and pathology of the fully annotated terminal 2

JOURNAL Mb of the short arm of human chromosome 16

MEDLINE Hum. Mol. Genet. 10 (4), 339-352 (2001)

PUBMED 11157797

REFERENCE 2 (bases 1 to 256073)

AUTHORS  
Daniels,R.J., Peden,J.F., Lloyd,C., Horsley,S.W., Clark,K.,  
Tufarelli,C., Kearney,L., Buckle,V.J., Doggett,N.A., Flint,J. and  
Higgs,D.R.

TITLE Direct Submission

JOURNAL Submitted (07-DEC-2000) MRC Molecular Haematology Unit, Weatherall  
Institute of Molecular Medicine, John Radcliffe Hospital, Oxford,  
Oxon OX3 9PS, UK

FEATURES Location/Qualifiers

source  
1. 256073  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16p13.3"  
1. 4641  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 356B8"  
3662. 46058

source

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 366d1"  
54069. 92001  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 338h10"  
76898. 117024

source

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 398g5"  
108489. 148123  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 349e10"  
159772. 192832

source

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 380a1"  
190568. 225840  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 444g9"  
219299. 256073

source

/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="cosmid 335h7"  
1. 3757  
/note="HS356B8"  
1. 747  
/note="Cpg island: JFPI/SOIH"  
/evidence=not\_experimental

misc\_feature

misc\_feature

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```

/note="PRAM; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
3658..39964
/note="HS366D1"
4193..4497
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
4507..4804
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
4805..5105
/note="AluY; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
5174..5490
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
complement(10481..10623)
/note="MER58; RepeatMasker predicted
23/10/2000"
/rpt_family="DNA/MER1 type"
complement(11229..11378)
/note="AluUd; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
complement(13829..14055)
/note="L1MB8; RepeatMasker predicted
23/10/2000"
complement(14418..14787)
/note="L1MB8; RepeatMasker predicted
23/10/2000"
/rpt_family="LINE/L1"
complement(14808..14991)
/note="AluUd; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
complement(15301..15507)
/note="AluXx; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
15303..15414
/note="AluUd; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
16935..17110
/note="(CCTCG)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
17340..17414
/note="AluUd; RepeatMasker predicted
23/10/2000"
/rpt_family="SINE/Alu"
join(19075..20523,21229..21437,21524..21707,21851..22074,
23338..23663,23748..23900,24287..24448,24537..24766,
24841..25007,25099..25277,25575..25752)
/gene="SOLH"
/note="PubMed:9722942"
/codon_start=1
/product="small optic lobes homolog"
/protein_id="AAK61233.1"
/db_xref="GI:1436701"
/db_xref="MIM:603267"
/db_xref="SPTREMBL:075808"
/translation="MATVGMSCVRCFLNPAGROCSICAPRHKPDNLHILRLVE

```

```

EOKMPCARCTFRNPLGKACEVCGTPEPAPGAFLFVLNGVLPKPAALIGEPKSCQ
EAGPVATGIVATTEPARGQCEBDEBEKEOEIEEAGAAPRGGMACPRCTLHNTVA
SSCGVGGPRRLSLRIPRALVPEVVAAGFVVAAPPGILPGGSAENPATSQ
GPAAREPPPPSPSTLONNVPVRSRREVPPOCPVPEAAGSPSPAGCGAPQ
GSGAGASRLAELISGRSLSVLEBRACGSTRVAGSSTSDIIDLAGTVYTPA
SPSPDFTTWSCACTLRNFTVAPRCACCGCKTHGQENHBPBTHDGDADKSPC
GRSGVSSAOKARVLPERPGWACACTILNLRKHCACAGTTPOLLVAKRGAP
LRRESNHEVQRQTDGEGAKLMENTVAFRENNVSPVDSFPGPSPVSPAGDSV
OQVRWMLRPOELNSVFRDHRATVYFHTLRPSDILGNCWFLSALVLERPD
LVERVWTRSLCAEAGAYOVLRCQDGTWTVTVDMLPCDEAGCLFESQAKROLVAL
TERKLATLHSYPLDQGRALBGLATLTGAPCESTALDSTNREBPVDLITAKM
LSSKEAGFLMGASCAGANMKVDASVSLGRPHNAYSILVROVQGRRLRLRNPKQ
RFSWNGSPSSADWPMPLRGELMPPHDSBGVFMWETGDFVRYFDSDVICKHSDMOE
ARVOCGSPSSAPAGVTALTILERRASLEFALPOGSRSDAVDSHLDLCTIVRAT
FGSGGHLSTGRLHSKRAVKEFVSCDMLERPEAVVCCAFNMHGPPLGTPAPQAS
SPKAGVPRASBPBPGHLYAVSSRLVWPEVPAOPTLADAILLITLBSGRHREGEG
MTCYTLTHGMAGLIVVENNRHPRKYLIVHQCCTSPNVYRSGSLRTODSVPLHROY
LVILISDQNGNAGFSITRHLARKAQAFPLSWTKSGHSPFLPEVAGLHGPRL"
19443..19488
/note="(GGA)n; RepeatMasker predicted
23/10/2000"
/rpt_family="Simple_repeat"
25280..25464
/note="C-rich; RepeatMasker predicted

Alignment Scores:
Pred. No.: 1,32e+03 Length: 256073
Score: 77.50 Matches: 15
Percent Similarity: 59.26% Conservative: 1
Best Local Similarity: 55.56% Mismatches: 10
Query Match: 45.86% Indels: 1
DB: 9 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x AB006464 (1-256073)
QY 4 GIYALAGLYAALALARY---GIYARGALASERTPYCTTPALALeulALeulTTP 22
DB 218086 GGGGCTGGGCTGGCCCGCAAGGGGCGGGGTGAGGTGGGCTGTGAGTCCGGTGG 218145
QY 23 LeuAlaValValPrGcLYTTP 29
DB 218146 ACGGATGACGCCCAAGTGG 218166

RESULT 15
AL137007 105779 bp DNA linear PRI 02-OCT-2000
LOCUS Human DNA sequence from clone RP1-30F19 on chromosome 6q12
DEFINITION Contains fragments of a gene similar to the Notch (Drosophila)
homolog, STS and GSBS, complete sequence.
ACCESSION AL137007
VERSION AL137007.9 GI:8745064
KEYWORDS HTG; Notch.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 105779)
REFERENCE
AUTHORS Williams,S.
TITLE Direct Substitution
JOURNAL Submitted (01-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
On Jun 26, 2000 this sequence version replaced gi:8573795.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence substitution
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the

```

sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
Emi; EMBL; Sw; SWISSPROT; Tr; TRMBL; Wp; WORMEP; Information on the WORMEP database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormep](http://www.sanger.ac.uk/Projects/C_elegans/wormep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Ch6>  
RP1-303F19 is from the library RP1-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>  
VECTOR: pCYPAC2  
This sequence is the entire insert of clone RP1-303F19. The true left end of clone RP1-401P4 is at 105680 in this sequence.

## FEATURES

## source

```
1.105779
/organism="Homo sapiens"
/mol_type="genomic DNA"
/dn_xref="taxon:9606"
/chromosome="6"
/map="q12"
/clone="RP1-303F19"
/clone_1b="RP1-1"
843..1047
/feature="MER63A repeat: matches 7..210 of consensus"
3043..3372
/feature="LI repeat: matches 3905..4233 of consensus"
3485..3623
/feature="LI repeat: matches 4226..4365 of consensus"
3624..3801
/feature="LIPB3 repeat: matches 5971..6146 of consensus"
3802..3815
/feature="LI repeat: matches 4365..4378 of consensus"
4413..4454
/feature="21 copies 2 mer ac 90% conserved"
5386..5777
/feature="MT1A1 repeat: matches 1..365 of consensus"
6598..6778
/feature="MER3 repeat: matches 1..209 of consensus"
8132..8221
/feature="MER1B repeat: matches 1..88 of consensus"
8222..8930
/feature="MER1A-internal repeat: matches 25..737 of consensus"
8916..9626
/feature="MER1A-internal repeat: matches 867..1580 of consensus"
9627..9979
/feature="MER1A repeat: matches 1..354 of consensus"
10218..10903
/feature="match: GSS: Em:AQ321138"
10422..10793
/feature="match: GSS: Em:AQ113885"
10808..11106
/feature="AluB repeat: matches 7..301 of consensus"
11321..11388
/feature="34 copies 2 mer to 67% conserved"
12656..13022
/feature="MER2CA repeat: matches 5..460 of consensus"
complement(12882..13318)
/feature="match: GSS: Em:AQ817672"
complement(12883..13322)
/feature="match: GSS: Em:AQ450753"
12884..13584
/feature="match: GSS: Em:B83082"
12928..13479
/feature="match: GSS: Em:AQ003116"
13093..13135
/feature="MER12CA repeat: matches 459..505 of consensus"
15274..15509
```

```
repeat_region
/feature="MER1 repeat: matches 310..563 of consensus"
17051..17264
/feature="MER1-INTERNAL repeat: matches 990..1211 of consensus"
17349..17779
/feature="LIP13 repeat: matches 5736..6156 of consensus"
17937..18338
/feature="MER2B repeat: matches 1..400 of consensus"
18369..18765
/feature="MER1C repeat: matches 65..462 of consensus"
complement(18807..19243)
/feature="match: GSS: Em:AQ034844"
19454..20541
/feature="LIP4 repeat: matches 5046..6146 of consensus"
20712..20890
/feature="LIMC repeat: matches 1682..1872 of consensus"
20964..21210
/feature="LIMC repeat: matches 1424..1676 of consensus"
21293..22072
/feature="LIMC repeat: matches 5489..6312 of consensus"
22102..22393
/feature="AluB repeat: matches 1..293 of consensus"
22394..22951
/feature="LIMC2 repeat: matches 4769..5328 of consensus"
22968..23066
/feature="LIM4 repeat: matches 2963..3065 of consensus"
23070..24449
/feature="LIM47 repeat: matches 4601..6003 of consensus"
24429..24982
/feature="LI repeat: matches 3787..4347 of consensus"
complement(24809..25281)
/feature="match: GSS: Em:AQ38604"
25033..25098
/feature="23 copies 2 mer ta 74% conserved"
25120..25259
/feature="LI repeat: matches 4377..4519 of consensus"
25260..25603
/feature="MER1B repeat: matches 1..364 of consensus"
25305..25905
/feature="match: GSS: Em:AQ491381"
25604..25690
/feature="LI repeat: matches 4518..4604 of consensus"
25806..25984
/feature="LIMC repeat: matches 2278..2099 of consensus"
25998..27419
/feature="MER52A repeat: matches 191..1755 of consensus"
27441..28042
/feature="LIMC repeat: matches 1672..2271 of consensus"
28570..29375
/feature="LIMC repeat: matches 944..1803 of consensus"
30178..30478
/feature="AluSC repeat: matches 1..303 of consensus"
30540..30562
/feature="Single clone region. Assembly confirmed by restriction digest"
30552..30838
/feature="AluSG repeat: matches 4..294 of consensus"
31345..31761
/feature="LI2 repeat: matches 1837..2274 of consensus"
33714..34154
/feature="match: GSS: Em:AQ114098"
35163..35945
/feature="LI2 repeat: matches 1837..2666 of consensus"
35947..36061
/feature="LIM4 repeat: matches 5615..5729 of consensus"
36069..36886
/feature="LIM4C repeat: matches 828..1627 of consensus"
38551..38643
/feature="LI2 repeat: matches 2631..2731 of consensus"
38845..39190
/feature="MER1B repeat: matches 1..364 of consensus"
39252..39633
/feature="LIM47 repeat: matches 5918..6288 of consensus"
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

4 protein - nucleic search, using frame\_plus\_p2n model

on: April 24, 2004, 23:02:32 ; Search time 4011.01 Seconds  
(without alignments)  
2352.639 Million cell updates/sec

File: US-10-010-050a-2\_COPY\_31\_346  
Effect score: 1758  
Sequence: 1 RVSGIPSRHRVDPYKRFDF.....IKTYEELPLIRNKLGL 316

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues  
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

MODE=frame+ p2n.model -DEV=x1h  
Q=/cpg2.1/USPTO.spool/US10010050/runat.22042004.113204.27556/app.query.fasta\_1.1372  
DB=EST -QFMT=fasta -SUFFIX=rev -MINMATCH=0.1 -LOOPEC=0 -LOOPEXT=0  
UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
USMR=US10010050 @CGN 1.1 6283 @runat.22042004.113204.27556 -NCPU=6 -ICPU=3  
NO\_MMAP -LARGEORDER -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: EST:\*

1:	em_estb1:
2:	em_estb2:
3:	em_estb3:
4:	em_estb4:
5:	em_estb5:
6:	em_estb6:
7:	em_estb7:
8:	em_estb8:
9:	em_estb9:
10:	em_estb10:
11:	em_estb11:
12:	em_estb12:
13:	em_estb13:
14:	em_estb14:
15:	em_estb15:
16:	em_estb16:
17:	em_estb17:
18:	em_estb18:
19:	em_estb19:
20:	em_estb20:
21:	em_estb21:
22:	em_estb22:
23:	em_estb23:
24:	em_estb24:
25:	em_estb25:
26:	em_estb26:
27:	em_estb27:
28:	em_estb28:

29: gb\_gss2.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1570	89.3	1201	9	AL546472
2	1423	80.9	2356	11	AK032247
3	1418	80.7	2247	11	AK032293
4	1415	80.5	2433	11	AK085741
5	1237	70.4	689	13	B0621797
6	1225.5	69.7	924	14	CA488543
7	1225	69.7	923	13	BX331615
8	1214	69.1	1201	9	AL571805
9	1211	68.9	931	10	B8873363
10	1191	68.9	869	12	BI090566
11	1167	66.4	908	13	BX370641
12	1159	65.9	974	13	BQ919385
13	1158	65.9	915	12	BI080303
14	1143	65.0	787	12	BI223533
15	1136	64.6	641	14	CB215456
16	1124.5	64.0	1077	13	BX342662
17	1121	63.8	938	13	BQ926170
18	1113	63.3	755	13	BQ443453
19	1076	61.2	1098	10	BF982158
20	1071	60.9	777	10	AW475730
21	1069	60.8	822	12	BI734549
22	1058.5	60.2	828	12	BI156421
23	1041.5	59.2	833	14	BI456284
24	1028	58.5	915	13	BK128221
25	1018.5	57.9	564	14	BK128221
26	1012	57.6	564	14	CB297292
27	969.5	55.1	676	12	BK57383
28	959	54.6	650	14	CF820113
29	953	54.2	680	14	CF820113
30	952	54.2	688	14	CF820113
31	938.5	53.4	767	12	CF820113
32	931	53.0	767	9	AL451053
33	922	52.4	869	14	CF378929
34	920.5	52.4	854	12	BI082103
35	901	51.3	607	12	BM849262
36	899	51.1	607	28	AQ309525
37	891.5	50.7	922	14	CA789305
38	882	50.2	661	13	BY734042
39	880	50.1	516	14	CB455600
40	857.5	48.8	736	12	BI077452
41	856	48.7	720	12	BI111605
42	843	48.0	623	13	BQ417448
43	841	47.8	604	10	BB622627
44	841	47.8	615	9	AL746877
45	834.5	47.5	706	10	BF739857

## ALIGNMENTS

RESULT 1  
AL546472  
LOCUS  
DEFINITION AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA  
clone CSOD1030Y01 5-PRIME, mRNA sequence.  
ACCESSION AL546472  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
EST.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 1201)

**AUTHORS** Li, M.-B., Gruber, C., Jesse, V. and Polyes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished (2001)  
**COMMENT** On Feb 15, 2001 this sequence version replaced gi:12879620.

**Contact:** Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 7238.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0D1030CE01QP1&cluster=7238.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue genoscope sequence ID : CS0D1030CE01QP1.

**FEATURES**  
 Location/Qualifiers

1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1030Y01"  
 /issue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_id="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-cligo (dt)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMSPORT 6 vector. Library was normalized."

# **ORIGIN**

## **Alignment Scores:**

**Pred. No.:** 4.01e-164 **Length:** 1201  
**Score:** 1570.00 **Matches:** 292  
**Percent Similarity:** 96.37% **Conservative:** 0  
**Best Local Similarity:** 96.37% **Mismatches:** 10  
**Query Match:** 89.31% **Indels:** 3  
**DB:** 9 **Gaps:** 1

US-10-010-050A-2\_COPY\_31\_346 (1-316) X AL546472 (1-1201)

QY 1 Argvalsergylleproserargarghistprrprovalprotyrlyarphasphpe 20  
 DB 153 CGGGTCCGGGATCCCTCCCGCGCCACTGCGCGGTGCTTACCAAGGCTTGACCTC 212  
 QY 21 Argprolyrproaspprotyrcysginalalstyrthphesysprothrglyserpro 40  
 DB 213 CGTCCAAACCTGATCTTATGTCAAGCTAAGTACTTTCTGCCAATGGCTCACC 272  
 QY 41 lleprovalmetgluylaspaaspaapillegluvalphearglueninalaprovalttrp 60  
 DB 273 ATCCGAGTTATGAGGGGTGATGATGACATTGAAGTTTTCGATTACAAAGCCCAATG 332  
 QY 61 Glupheryltyrlylaspleuleuglyhisleulyllemerhisapalalegylphe 80  
 DB 333 GAATTAAATATGAGACCTCCCGGACACTGAAATATATGATATGATGATGATGATG 392  
 QY 81 Argserthreuthrglylsasenytyrthmetglutryrlygluendpneugly 100  
 DB 393 AGAAGTACATTACTGCGAAGACTACACATGGAATGATGATGATGATGATGATGATG 452  
 QY 101 Asncysrthreuthrglylsasenytyrthmetglutryrlygluendpneugly 120  
 DB 453 AATGTGACATTTCCCATCTCCGACCTGAAATGATGATGATGATGATGATGATG 512  
 QY 121 Alalalacysphesphesgluylleaspaaspaalhistrlyrlygluendpneugly 140  
 DB 513 GCTGCCGCTTTTGTGGGAATGATGATGATGATGATGATGATGATGATGATGATG 572  
 QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleStrpValIlysglnasp 60  
 DB 573 CAAATGCAACCTATATAGGAACATGTTCAACCAATGGCAAGTGGGGAACGAGAC 632  
 QY 161 AsnGluThrIleGlyIleTyrrgluThrTrpAsnValIlysalaserprogluylgllyala 80

DB 633 AATGAAACGGAATTTATTTATATGACATGATGATGATGATGATGATGATGATGATG 692  
 QY 181 GluThrTrpPheAspSerTyraapCyserlyspheValleuargThreAsnlyleu 200  
 DB 693 GAGACATGGTTGATCTCAGACCTGCTTCAATTTGTGTAAGGACCTTTTACCAAGTGG 752  
 QY 201 AlaGluPheGlyAlaGluPheAsnleleGluThrAsnTyrrThrglylePheleuTy 220  
 DB 753 GCTGAATTTGAGACAGATGTCAGAACATGAAACCACTATACAAATATTTCTTTAC 812  
 QY 221 SerGlyGluProthrrtyrlyrlyleuglylaenGluThrSerValPheGlyProThrglyAsnly 240  
 DB 813 AGTGAAGAACCTTACTTATCTGGAAATGAAACATCTGTTTTCGGGCAACAGGAACAG 872  
 QY 241 ThrLeuGlyLeuAlaIleIysArgPheTyrrTyrrProPheAspProHisleuProThrl 260  
 DB 873 ACCTTGTTTACCAATTAATAAATTTTATTTACCCCTTCAACACATTTGCCAATCA 932  
 QY 261 GlupheleuleuSerleuleuGlnIlePheAspAlaValleValHisIlysglnPheTy 280  
 DB 933 GAATTTCTGTGAGTCTCTTGAATTTTGAATGATGATGATGATGATGATGATGATG 992  
 QY 281 LeuphetTyraasphesgluTyrrTyrrPheleuProMetIysPheProPheIleTy 300  
 DB 993 TTGTTTATTAATTTTGA-TATGGTTTTTMM---CYATGAAMTCCCTTTATTAAMA-ACA 1047  
 QY 301 TyrrgluGlu 303  
 DB 1048 TATGAGGAA 1056

**RESULT 2**  
 AK043247 2356 bp mRNA linear HTC 19-SEP-2003  
 LOCUS AK043247.1 GI:26089577  
 DEFINITION Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length  
 enriched library, clone:A730075N08 product:similar to  
 CEROID-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLIN PROTEIN) [Homo  
 sapiens], full insert sequence.  
 ACCESSION AK043247  
 VERSION AK043247.1 GI:26089577  
 KEYWORDS HTC; CAP trapper.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

**REFERENCE**  
 1 Carninci, P. and Hayashizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999)  
 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Kono, H., Akiyama, Y., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Kashiwagi, K.,  
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Waburata, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 Genome Res. 10 (11), 1757-1771 (2000)  
**JOURNAL** MEDLINE 20530913  
**PUBMED** 11076861  
**REFERENCE** 4

**AUTHORS** The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.  
**TITLE** Functional annotation of a full-length mouse cDNA collection  
**JOURNAL REFERENCE** Nature 409, 685-690 (2001)  
**AUTHORS** 5  
**TITLE** The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
**JOURNAL REFERENCE** Analyses of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
 Nature 420, 563-573 (2002)  
**AUTHORS** 6 (bases 1 to 2356)  
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komori, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takada, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Direct Subsession  
**JOURNAL REFERENCE** Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
**COMMENT** cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN. Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: http://genome.gsc.riken.go.jp/  
 Location: Qualifiers  
**FEATURES**  
 source  
 1. 2356  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:A730075N08"  
 /db\_xref="MGI:2408405"  
 /db\_xref="taxon:10090"  
 /clone="A730075N08"  
 /tissue\_type="cerebellum"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="7 days neonate"  
 1. 2356  
 /note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CIN5 PROTEIN) [Homo sapiens] (SWISSPROT:O75503, evidence: FASTA, 73.9%ID, 86.7%length, match=1041)"  
**ALIGNMENT SCORES:**  
 red. No.: 2.28e-147 Length: 2356  
 core: 1423.00 Matches: 246  
 percent similarity: 89.03% Conservative: 30  
 percent local similarity: 79.35% Mismatches: 34  
 query match: 80.94% Indels: 0  
 B: 11 Gaps: 0  
 US-10-010-050a-2\_COPY\_31\_346 (1-316) x AK043247 (1-2356)  
 Y 7 SetAArgArgHisTTPProValPProTyrTyrAlaGpheaAphAeAProLysProAaspPro 26  
 b 148 TCTGGGCAAGCGTGGCGGCGCTTCTCTTCTCGTCCGAGACAGATCC 207  
 Y 27 TycGagGlnAlaLysTyrThrPheCysProThGlySerProIleProValMetGluGly 46

Db 208 TACTGCAAGCTAAGTAACTTTCTGTCTACCGGCTCGCCCATCCAGTTATGAGAGAC 267  
 Qy 47 AspaApepIleGluValPheaArgLeuGlnAlaProValTPGluPheLysTyrGlyAasp 66  
 Db 268 AATGACGTCATCGAGGCTTACGACTCAAGACCCCGGATTTGAGATTTAAATATGAGAC 327  
 Qy 67 LeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGly 86  
 Db 328 CTCCTGGACACTTTAAACTTATGTCATACCCCGGAGATTCAGAGACACATCGACAGGC 387  
 Qy 87 LysAsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHis 106  
 Db 388 AAGAATCAACATTAAGATGATGATGAATTTTCCAGCTGGGCACTGATCAATTCGCCAC 447  
 Qy 107 LeuArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGlu 126  
 Db 448 CTCGGCGCTCGCAAGAGCGCTCCCTCTCGGTGATCAAGGGGCGACCTGCTTTTGTGA 507  
 Qy 127 GlyIleAspApepValHisTyrPheGluAsnGlyThrLeuValGlnValAlaThrIleSer 146  
 Db 508 GGAATGATGATAAACACTGGAAGAAAACGGACACTGTCACTGCTTGCAACCATATCC 567  
 Qy 147 GlyAsnMetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluThrGlyIleTyr 166  
 Db 568 GGAACACATTTTAAACAAGTGGCCGAGTGGGGAAGCAGACATGAATGAGATTAT 627  
 Qy 167 TyrGluThrTPAsnValLysAlaSerProGluLysGlyAlaGluThrTPheApeSer 186  
 Db 628 TATGAGACATGACCGTCCGACCGCCGACGAGCAAGGGCCCAACGCTGTTTCAGTCC 687  
 Qy 187 TyrApeCysSerLysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGlu 206  
 Db 688 TACGACTGTTCGAATTTGTCTTAAGACATATAGAAATTTGGCTGAATTTGGAACAGA 747  
 Qy 207 PheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyr 226  
 Db 748 TTCAGAAGATGAGAAACAACTATACGAAATATTTCTTACAGTGGAGAGCTATTAC 807  
 Qy 227 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle 246  
 Db 808 CTGGGAATATGAAACATCTATTTTGGGCCCAAGAAAACAGACTTCTTGCCATA 867  
 Qy 247 LysArgPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuSerLeu 266  
 Db 868 AAAAATTTTATAGCCCTTCAGACCGATTTGTCAACCAAGATTTCTGATGATTTTC 927  
 Qy 267 LeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGlu 286  
 Db 928 TTGAATAATTTTGAATACAGTATATACACAGACAGTCTTACTGTGTTATATACCTTTCAG 987  
 Qy 287 TyrTyrPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeu 306  
 Db 988 TATTTGTTTCTTCAATGAAACCCCTTGTTCAAATATACATGACAAAGAAACCCGTTA 1047  
 Qy 307 ProIleArgAsnLysThrLeuSerGlyLeu 316  
 Db 1048 CCAACCCGACATACACATTTTACCACTTG 1077  
**RESULT 3**  
 AK032293 2247 bp mRNA linear HTC 18-SEP-2003  
**LOCUS**  
**DEFINITION** Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430516P20 product:similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CIN5 PROTEIN) [Homo sapiens], full insert sequence.  
**ACCESSION** AK032293  
**VERSION** AK032293.1 GI:26328112  
**KEYWORDS** HTC; CAP trapper.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
**REFERENCE** 1  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE  
AUTHORS 2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE  
AUTHORS 3  
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishii, K., Kiteunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasahara, A.,  
Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multichannel sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE  
AUTHORS 4  
The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
TITLE Functional annotation of a full-length mouse cDNA collection  
JOURNAL Nature 409, 685-690 (2001)  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS 5  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
TITLE Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL Nature 420, 563-573 (2002)  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS 6  
(bases 1 to 2247)  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirose, T.,  
Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, K.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishii, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takai-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
TITLE Direct Submission  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsunumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gscl.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
COMMENT  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/.  
FEATURES  
source  
1..2247  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="PANTOM DB:6430516p20"  
/db\_xref="WGI:2335762"

ORIGIN  
Alignment Scores:  
Pred. No.: 7,77e-147 Length: 2247  
Score: 1418.00 Matches: 245  
Percent Similarity: 86.71% Conservative: 30  
Best Local Similarity: 79.03% Mismatches: 35  
Query Match: 80.66% Indels: 0  
DB: 11 Gaps: 0  
US-10-010-050a-2\_copy\_31\_346 (1-316) x AK032293 (1-2247)  
QY 7 SerArgArgHisIstProValProIlylSAyGpPheAspPheArgProIysProAspPro 26  
||| :|||  
Db 39 TCTGGGACGACGCTGCGCGGCGCATCAAGCACCCTCTCTCTCGTCCGAGACAGANTCC 98  
27 TyCyGGLAlAlayrTyThrPheCySProThGlySerProIleProValMetGluGly 46  
||| :|||  
Db 99 TACTGTCAAGCTAAGTACTTCTGCTTCTGCTTCCGCGCTCCCATCCAGTTAAGAAGAC 158  
47 AspAspAspIleGluValPheArgLeuGlnAlaProValIrpGluPheIySTyGlyAsp 66  
||| :|||  
Db 159 AATGACGCTATCGACAGCTTACGACATCAAGCCCGATTGGGAATTAATGAGAC 218  
67 LeuLeuGlyHisLeuIySileMeChIaSPAlaIleGlyPheArgSerTrIleuThGly 86  
||| :|||  
Db 219 CTCTGGGACACTTAACTTATGATGACGCGCGGATTCAGGACACATGACAGGC 278  
87 LysAsnTyThrMetGluTyTyTyGluLeuPheGlnLeuGlyAsnCyThrPheProHis 106  
||| :|||  
Db 279 AAGAACTACCAATGAGTGTGATGAATTTTCCAGCTGGGCAATGTACATTCCCCAC 338  
107 LeuArgProGluMetAspAlaProPheTyCyAsnGlnGlyAlaAlaCySPhePheGlu 126  
||| :|||  
Db 339 CTCGGGCTTCAAGACAGCGCTCCCTTCTGTGTGATCAAGAGGCGCAGCTCTTTTGAA 398  
127 GlyIleAspAspValHisIstIrylGlyAsnGlyThrIleuValGlnValAlaThrIleSer 146  
||| :|||  
Db 399 GGAATAGATATATAACCTGGAAGAAACGGGACATGTCAGTGTGGAACCATATCC 458  
147 GlyAsnMetPheAsnGlnMetAlaIyEtrValIyGlnAspAsnGluThrGlyIleTy 166  
||| :|||  
Db 459 GGAACACATTTTAAACAAGTGGCCGAGTGGTGAAGAGACAAATGAATCGGATTTAT 518  
167 TyGluThrTyPheAsnValIyAlaSerProGluIyGlyAlaGluThrTyPheAspSer 186  
||| :|||  
Db 519 TATGAGACATGAGAGGTCGCGCGCGCCGACGACAGAGGCGCCGAGTGTGAGTCC 578  
187 TyTAspCySerIySPheValIleuArgThyPheAsnIySileuAlaGluPheGlyAlaGlu 206  
||| :|||  
Db 579 TACACAGTTCGAATTTTGTCTTAAGACATATGAATTTGGCTGAATTTGGAACAGAA 638  
207 PheIyAsnIleGluThrAsnTyThrArgIlePheLeuTySerGlyGluProThrTy 226

```

|||||
639 TTCAAGAAAGATAGAAACAACTATACGAAATATTTCTTTCAGTGGAGACCTATTTAC 698
227 LeuGlyAangUtuThrservAlPheGlyProthrglyAenUsthrLeuGlyLeuAlle 246
699 CTGGAAATGAAACCTCATTTTGTGGGCCCAAGAAACAAAGACTTGTGCTTGGCCATA 758
247 LysArgPheTyTyTyProPheIysProHisIleuProthrglyGluPheLeuIleuSerIleu 266
759 AAAAAATTTTATGCGCCCTTCAGACCGATTTGTCAACCAAGATTTTCTGATGAAATTC 818
267 LeuGlnIlePheAspAlaValIleValHisIlyseGlnPheTyTyLeuPheTyTyAspPheGlu 286
819 TTGAAATTTTGTATACAGTGTATATACACAGACAGTCTACTGTTTATTAACCTTGAG 876
287 TyTyTrpPheIleuProMetIysPhePheProHeliIeTyTyTyGluGluIleProIleu 306
879 TATGTTGTTTCAACATGAAACCCCTTTGTCAAAATTAACATGACAAAGAACCCCGTTA 936
307 ProIleArgAsnUtyThrIleuSerGlyLeu 316
939 CCTACCCACATACACATTTTACGACTTG 968

RESULT 4
1085741 2433 bp mRNA linear HTC 20-SEP-2003
XOUS Mus musculus 10 days lactation, adult female mammary gland cDNA,
PRINTITION RIKEN full-length enriched library, clone:D730033P03
PRODUCT: similar to CEROID-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5
PROTEIN) [Homo sapiens], full insert sequence.
XSESSION AK085741
XRSION AK085741.1 GI:26102938
XWORDS HTC; CAP trapper.
XORCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
JOURNAL High-efficiency full-length cDNA cloning
MEDLINE Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
TITLE Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalisation and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
20499374
MEDLINE 11042159
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kikunishi, T., Tachino, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Hara, A.,
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
20530913
MEDLINE 11076861
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
TITLE FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
5
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
TITLE Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation

```

```

JOURNAL of 60,770 full-length cDNAs
REFERENCE Nature 420, 563-573 (2002)
6 (bases 1 to 2433)
AUTHORS Aachidi, J., Aizawa, K., Akimura, T., Arakawa, T., Boro, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiroaka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, K., Onno, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toy, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
source
1. 2433
location/qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:D730033P03"
/db_xref="MG1:2422937"
/db_xref="taxon:10090"
/clone="D730033P03"
/sex="female"
/tissue_type="mammary gland"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days lactation, adult"
1. 2433
/note="similar to CEROID-LIPOFUSCINOSIS NEURONAL PROTEIN 5
(CLN5 PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence:
FASTV, 73.3%ID, 86.7%length, match=1041)"
ORIGIN
misc_feature
1. 85e-146 Length: 2433
Pred. No.: 1415.00 Matches: 245
Score: 88.71% Conservative: 30
Percent Similarity: 79.03% Mismatches: 35
Beet Local Similarity: 80.43% Indels: 0
Query Match: 11 Gaps: 0
DB:
US-10-010-050A-2_COPY_31_346 (1-316) x AK085741 (1-2433)
QY 7 sepArgAGhIstTTPProValProTyTyIysAsgPheAspPheArgProIysProAspPro 26
225 TCTGGGACAGCGTGGCCGCTGCTTCAACGCGCTTCTTTCGCTCGAAGACAGATCC 284
QY 27 TyTyGlnAlaIlyTyTyThrPheTyProthrglySerProIleProValMetGluGly 46
285 TACTGCAAGCTAAGATATCTTCTGCTTCAACGCGCTCGCCATCCAGTTATGAAGAC 344
QY 47 AspAspAspIleIleValPheArgLeuGlnAlaProValTrpGluPheTyTyGlyAsp 66
345 AATGAGGTGATTCAGAGCTTACGACTACAAAGCCCAATTTTGAATATGAGAC 404
QY 67 LeuLeuGlyHisIleuIysIleuMetHisAspAlaIleGlyPheArgSerThrIleuThrgly 86

```

```

|||||
b 405 CTCTGGGACACTTAACTTATGACATGACCGCGTGGGATTTCAGAGCACACATGACAGGC 464
y 87 LysAsnTyrThrMetGluTrrPylGluLeuPheGlnLeuGlyAsnGlyThrPheProHis 106
b 465 AAGAACTACACATGATGAGTGAAGACCTTTCAGCTGGGCACTGTACATTTCCCCAC 524
y 107 LeuArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaIaCysPhePheGlu 126
b 525 CTCCGGCTTACACAGAGCGCTCCCTTCTGTGTATGACCAAGGGCAGCTGCTTTTGA 584
y 127 GlyLeuAspValHisIstrPylGluAsnGlyThrLeuValGlnValAlaThrIleSer 146
b 585 GGAATATGATGATTAACACTGGAAGAAACGGGACACTGTCACTGTGTGCAACCATATCC 644
y 147 GlyAsnMetPheAsnGlnMetAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 166
b 645 GGAACACATTTACCAAGTGGCGAGTGGGTGAAGACAGACATGAACTGGGATTTAT 704
y 167 TyGluThrTrpAsnValIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 186
b 705 TATGACATGACATGACATGACATGACATGACATGACATGACATGACATGACATGAC 764
y 187 TyAspCysSerIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 206
b 765 TAGACATGCTGATTTGCTTCTTAAAGACATATTAAGAAATTTGCTGATTTGGAACGAA 824
y 207 PheIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 226
b 825 TTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 864
y 227 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnGlyThrLeuGlyIle 246
b 885 CTGGGAAATGAACATCTATTTTGGGCGCAAGAAAGAAACAACTCTTGGCTTGGCATA 944
y 247 LysArgPheTyrTyrProPheIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 266
b 945 AAAAATTTTATGGCCCTTCCAGACCGATTTTGCACCAAGAAATTTTCTGATTAATTC 1004
y 267 LeuGlnIlePheAspAlaValIleValHisIaIaIaIaIaIaIaIaIaIaIaIaIa 286
b 1005 TTGAAATTTTGTATACAGTATATACAGACAGAGCTTCTTCTTTTATTAACCTTGAG 1064
y 287 TyTrpPheLeuPhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 306
b 1065 TATGCTTTCTTACCAATGAAGAACCCCTTGTCAAAATTAATGAAGAAACCCCTTGA 1124
y 307 ProIleArgAsnIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 316
b 1125 CTTACCCGACATACAAACATTTACCGACTTG 1154

```

```

RESULT 5
BU621797/c 689 bp mRNA linear EST 23-SEP-2002
LOCUS BU621797
DEFINITION UI-H-FL1-b9b-i-07-0-UI-b1 NCI CGAP FL1 Homo sapiens cDNA clone
BU621797
ACCESSION BU621797.1 GI:23288012
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 689)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT Contact: Robert Strausberg, Ph.D.
E-MAIL Email: cgaabs-remail.nih.gov
TISSE Tissue Procurement: James Martin
CDNA CDNA library preparation: Dr. M. Bento Soares, University of Iowa
CDNA CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa

```

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@iowa.edu  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

## FEATURES

source

1. 689 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FL1-b9b-i-07-0-UI"  
 /tissue\_type="Cell lines"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NCI CGAP FL1"  
 /note="Organ: Chondrosarcoma; Vector: pT73-Pac  
 (Pharmacia) with a modified polylinker; Site 1: EcoR I,  
 Site 2: Not I; NCI CGAP FL1 is a normalized cDNA library  
 derived from a pool of mRNA obtained from 4 cell lines  
 from grade III chondrosarcoma tissues. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA,  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor digested with Not I, and cloned directionally  
 into pT73-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 GAGGTCGGTG. The cell lines were provided by Dr. James  
 Martin from the University of Iowa.  
 TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix  
 TAG\_LIB=UI-H-FL1  
 TAG\_SEQ=GAGGTCGGTG"

## ORIGIN

## Alignment Scores:

Pred. No.: 2 71e-127 Length: 689  
 Score: 1237.00 Matches: 222  
 Percent Similarity: 100.00% Conservative: 3  
 Best Local Similarity: 98.67% Mismatches: 0  
 Query Match: 70.36% Indels: 0  
 DB: 13 Gaps: 0

US-10-010-050a-2\_copy\_31\_346 (1-316) x BU621797 (1-689)

```

Oy 79 GlyPheArgSerThrLeuThrGlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 98
Db 688 GGAATTCAGAGATACATTAATGCGGACAGACATGACATGATGATGAACTTTCCAA 629
Oy 99 LeuGlyAsnGlyThrPheProHisIleuArgProGluMetAspAlaProPheTyrCysAsn 118
Db 628 CTGGCAACTGTATACATTTCCCATCTCCGACCTGAATGAGATGCCCTTTCTGTGTAT 569
Oy 119 GlnGlyAlaIaIaCysPhePheGlnGlyIleAspAspValHisIstrPylGluAsnGlyThr 138
Db 568 CAAAGCGCTGCTCTCTTTTGGAGGAATGATGATGCTCTGGAAGAAATGAGGACA 509
Oy 139 LeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIaIaIaIaIa 158
Db 508 TTAGTTCAAGTAGACATATATACAGAAACATGTTCAACCAATGGCAAAATGGGTGAAA 449
Oy 159 GlnAspAsnGluThrGlyIleTyrTyrGluThrTrpAsnValIaIaIaIaIaIaIaIa 178
Db 448 CAGACATATGAAGAGAGATTTTGGAGCAATGAGATGTAAGCCAGCCGAGAAAG 389
Oy 179 GlyIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 198
Db 388 GGGCAGAGAGATGTTGATTCCTACAGACTGTTCCAAATTTGTTAAGACCTTTTAC 329
Oy 199 LysLeuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 218
Db 328 AAGTTGCTGAATTTGGAGCAGAGTTCAAGAAACATGAAACCACTATACAAATATTT 269

```



219 leuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 2:8  
 268 CTTACAGTGGAGAACTACTATTGGGAATGAACATCTGTTTGGCCAAACGGA 209  
 239 AsnTyrThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuPro 2:58  
 208 AACAAAGACTCTGGTGGTCCATAAAAGATTATTATACCCCTCAACACCATTTGGCA 149  
 259 ThrTyrGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGln 2:78  
 148 ACTAAAGATTCTGTTAGTCTCTGCAAAATTTTGATGAGATTTGTCAACAAACG 89  
 279 PheTyrLeuPheTyrAsnPheGluTyrTyrTyrPheLeuProMetLysPheProPheIleLys 2:98  
 88 TTTCAATTTGTTTATATTTGAAATATGTTTACTATGAATATCCCTTTATTATAA 23  
 299 IleThrTyrGluGln 303  
 28 ATTAACATTGAAAGA 14

RESULT 6  
 CA488543  
 LOCUS  
 DEFINITION AGENCOURT 10808884 MAPcl Homo sapiens cDNA clone IMAGE:6720309 5',  
 mRNA sequence.  
 ACCESSION CA488543  
 VERSION CA488543.1 GI:24950702  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE  
 1 (baee 1 to 924)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsabers@mail.nih.gov  
 Tissue Procurement: Kristi A. Egland, Ira Pastan  
 cDNA library preparation: Invitrogen Corp  
 cDNA library arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM14279 row: j column: 21  
 High quality sequence stop: 637.

## FEATURES

source

1..924  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6720309"  
 /cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,  
 hTERT-HMEL, LNCaP"  
 /lab\_host="EMDHL08"  
 /clone\_1ib="MAPcl"  
 /note="Vector: PCMV-SPORE6; Site 1: EcoRV; Site 2: Not I;  
 Subtracted with brain, liver, lung, kidney and muscle.  
 Directionally cloned. Priming method: oligo-dt. Average  
 insert size: 1800 bp. Library amplification: 26,000 fold.  
 Kristi A. Egland, James J. Vincent, Robert Strausberg,  
 Bungkok Lee & Ira Pastan. Discovery of new breast  
 cancer genes encoding membrane and secreted proteins.  
 Manuscript submitted."

## ORIGIN

Alignment Scores:  
 Pred. No.: 7,48e-126 Length: 924  
 Score: 1225.50 Matches: 229  
 Percent Similarity: 87.73% Conservative: 7  
 Best Local Similarity: 85.13% Mismatches: 20

Query Match: 69.71% Indels: 13  
 DB: 14 Gaps: 3  
 US-10-010-050a-2\_copy\_31\_346 (1-316) x CA488543 (1-924)

QY 1 ArgValSerGlyIleProSerArgArgHisIleTyrProValProTyrIlyArgPheAspPhe 20  
 DB 106 CGAGTCTCGGGGCACTCCCTCCGGGCGGCACTGGCGGGTGCCTTCAAGCGCTTTGACTTC 165  
 QY 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40  
 DB 166 CGTCAAAACCTGATCTCTTATGTCAGCTAAGTATCTTCTGTCCAACTGGCTCAGCT 225  
 QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60  
 DB 226 ATCCAGATTAGAGGGGTGATGATGACCTTGAAGTTTTCATTTCAATTAAGCCCACTAGG 285  
 QY 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 80  
 DB 286 GAATTTAATATGAGAGACTCTCTGGGACACTGAAATTAATGATGATGATGATGATGATG 345  
 QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGluLeuPheGluLeuGly 100  
 DB 346 AGAATGATCATTAATCTGGCAAGAACTACCAATGAAATGATGATGATGATGATGATGATG 405  
 QY 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120  
 DB 406 AACTGTACATTTCCCACTCCGACCTGCAATGAAATGATGATGATGATGATGATGATGATG 465  
 QY 121 AlaAlaCysPhePheGluGlyIleAspAspValHisTyrPylGluAsnGlyThrLeuVal 140  
 DB 466 GCTGCTCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 525  
 QY 141 GluValAlaThrIleSerGlyAsnMetLeuGlnMetAlaLysTyrValIleGlnAsp 160  
 DB 526 CAAGTAGCACTATATGCAAGAAACATGTTCACCAAAAGGCAATGGGTGTAACAGGAGC 585  
 QY 161 AsnGluThrGlyIleTyrTyrGluThrTyrPheValIleValAspProGluLysGlyAla 180  
 DB 586 AATGAAACAGAAATTTATATGAGACATGAGATGTAAAGCCAGCCAGAAAGGGGGCA 645  
 QY 181 GluThrTyrPheAspSerTyrAspCysSerLysPheValIleuArgThrPheAsnLysLeu 200  
 DB 646 GAGACATGTTGATTTCTTCAAGACCTGTCAAATTTGTGTTAGACCTTTAAACGGTTG 705  
 QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArg-IlePheLeuT 220  
 DB 706 GCTGAATTTGGAGCAGAGTTCAAGAACTAGAAACCAATATACAGAAATATTTCTTT 765  
 QY 220 YrSerGlyGluPro---ThrTyrLeuGlyAsnGluThrSer-----ValPheGly---- 235  
 DB 766 ACAGTGGGAGAAACCTTACTTATCCGGGGAATTTGAAAACATTCGATTTTGGGGCCT 825  
 QY 236 -----ProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgP 249  
 DB 826 AACAGAGAAACCAAGAACTCTTGGGTGTTTACCCCATTAACCAAAAAAATTTTAAAT 885  
 QY 249 heTyrTyrProPheLysProHis 256  
 DB 886 TAACCGCCCTTCTCAAAACAC 908

RESULT 7  
 EX331615  
 LOCUS  
 DEFINITION EX331615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens  
 cDNA clone CS00B009YC01 5-PRIME, mRNA sequence.

ACCESSION EX331615.1 GI:30310073  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 923)  
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DB009AB01QP1&cluster=7238.f. Contact :  
Peng Liang Email: fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DB009AB01QP1.

FEATURES  
source

1..923  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DB009Y01"  
/issue\_type="NEUROBLASTOMA COT 10-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.:	8,49e-126	Length:	923
Score:	1225.00	Matches:	221
Percent Similarity:	93.64%	Conservative:	0
Best Local Similarity:	93.64%	Mismatches:	15
Query Match:	69.68%	Indels:	0
		Gaps:	0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x BX331615 (1-923)

```

QY 1 ArgValSerGlyIleProSerArgArgHisTrpProValProTyrLysArgPheAspPhe 20
Db 208 CGGGCTCGGGCAGTCCCTCCCGGCGCCACGCGCGGCTTCAACCGCTTGACTTC 267
QY 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
Db 268 CGTCCAAAACCGATCCCTTATGTGACGTAAAGTATACCTTCTGTCCAACTGGCTCACT 327
QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp 60
Db 328 ATCCCACTTATGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 387
QY 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysValIleMetHisAspAlaIleGlyPhe 80
Db 388 GAATTTAAATATGAGACCTCTCTGGACACTTGAATAATATATGATGATGATGATGATGAT 447
QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 100
Db 448 AGAAGTACATTACCTGGAGAGAACTAATACATGATGATGATGATGATGATGATGATGATG 507
QY 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120
Db 508 AANTGTAAATTTCCCTCCCTCCGNNCTGNAATGAGTSSCTTCTGGGATATCAAGG 567
QY 121 AlaAlaCysPhePheGluGlyIleAspAspValHisTrpLysGlnAsnGlyThrLeuVal 140
Db 568 GSTGCTGCTTTTTRGGGAATGTGTGATGTTCACTGGAAGGAAATGGACATTAGTT 627
QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrValIlysglnAsp 160
Db 628 CAAGTGCAACTATATCAGAAACATGTTCAACCAATGGCAAGAGGGGTGAAAAGAGAC 687
QY 161 AsnGluThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180

```

```

Db 688 AATGAAACGGAATTTATTATGAGCATGATGATGAAAGCCAGCAAGAGGGGCA 747
QY 181 GluThrTrpPheAspSerTyrAspCysSerLysPheValIleArgThrPheAsnLysLeu 200
Db 748 GAGACATGTTGATTTCTACGACATGTTCCAAATTTGTGTTAAGGACCTTTAAACAAGTTG 807
QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
Db 808 GGTGATTTGGAGCGAGTTCAAGAACATAGAAACCACTATCAAGATATTTCTTAC 867
QY 221 SerGlyLysProThrTyrLeuGlyAsnGluThrSerValPheGlyPro 236
Db 868 ATGCGGAACCTATCTATCTGGGAATGAAACATCTGTTTGGGCCA 915

```

RESULT 8  
AL571805/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AL571805 1201 bp mRNA linear EST 31-MAY-2003  
AL571805 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1030Y01 3-PRIME, mRNA sequence.  
AL571805.2 GI:31293196  
EST.  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On Feb 16, 2001 this sequence version replaced gi:12929467.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f. Contact :  
Peng Liang Email: fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0D1030CE01NP1.

COMMENT

JOURNAL

TITLE

AUTHORS

REFERENCE

FEATURES  
source

1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1030Y01"  
/issue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.:	1.99e-124	Length:	1201
Score:	1214.00 <td>Matches:</td> <td>230</td>	Matches:	230
Percent Similarity:	96.64%	Conservative:	0
Best Local Similarity:	96.64%	Mismatches:	7
Query Match:	69.06%	Indels:	3
		Gaps:	0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AL571805 (1-1201)

```

QY 80 PheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeu 99
Db 1067 TTCAGAGTACATTACCTGGGCAATG-TACAAATGATGATGATGATGATGATGATGATG 1010
QY 100 GlyAsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGln 119

```

```

1009 GGCACACTGATATTCCTCCGACCTGAAATGATSCCCCTTCGTGATGATCAA 950
120 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGAT 139
949 GGGCTGCTGCTGCTTTTGGAGGAAATGATGATGATGATGATGATGATGATGAT 830
139 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 139
889 AGTTCAGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 830
159 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 179
829 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 770
179 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 199
769 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 710
159 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 219
709 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 650
219 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 239
649 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 590
239 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 259
589 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 530
259 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 279
529 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 470
279 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 299
469 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 410
299 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 316
409 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 358

RESULT 9
BE873363 931 bp mRNA linear EST 20-OCT-2000
LOCUS BE873363 601450425F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3854119 5',
DEFINITION mRNA sequence.
ACCESSION BE873363
VERSION BE873363.1 GI:10322139
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strauberg, Ph.D.
Email: c9apbs-f@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L14M9579 row: b column: 08
High quality sequence stop: 662.
Location/Qualifiers
1. 931
SOURCE /organism="Homo sapiens"

```

```

/mot_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3854119"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_id="NIH_MGC_65"
/note="Organ: colon; Vector: pCMV-Sport6; Site: 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

ORIGIN

Alignment Scores:
Pred. No.: 3.12e-124 Length: 931
Score: 1211.00 Matches: 232
Percent Similarity: 93.57% Conservative: 1
Best Local Similarity: 93.17% Mismatches: 14
Query Match: 68.89% Indels: 4
DB: Gaps: 1

US-10-010-050a-2_copy_31_346 (1-316) x BE873363 (1-931)

QY 47 AASPAPAPILLEGUVALPHEARGLEUINALAPROVALTPGJUPHELYSTYGLYASP 66
DB 2 GATGATGACATTTGATGAGGAAATGATGATGATGATGATGATGATGATGATGAT 61
QY 67 LEULEUGLYHILEULYVILEMECHSAASPALILEGYPHEARGSERTHLEUTHG 86
DB 62 CTCCTGGACACTTGAAATATGATGATGATGATGATGATGATGATGATGATGAT 121
QY 87 LYSAANTYRTHMETGLUTRTPYRGLUUPHEGILEUGLYASNCYETHRPHROHIS 106
DB 122 AAGAATCTACAAATGAAATGATGATGATGATGATGATGATGATGATGATGAT 181
QY 107 LEUARGPROGJUMETASPLAPROBHEITRPSAENGINGLYALALACYPHEPHE 126
DB 182 CTCGACCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
QY 127 GYILLEASPAPVALHISTRLYGLUENGILYTHRLUVALGLVALATHRIEGER 146
DB 242 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
QY 147 GLYASNMETPHEASNGIMETALALYSTPVALYSGINASPASNGIUTHRGYILE 166
DB 302 GGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361
QY 167 TYRGJUTHRTPASNVALLYSAIASERPROGJULYGLYALGLUUTHRTPHEAS 186
DB 362 TATGAGACATGAAATGATAAAGCCAGCCAGAAAGGGGACAGACATGTTGATTC 421
QY 187 TYRASPYSERLYPHEVALLEUARGTHRPHASNVLEUALGLUPHEGLYALGLU 206
DB 422 TACGACTGTTCCAAATTTGTTAAGACCTTTAACAAATTTGCTGAATTTGGAG 481
QY 207 PHELYASNVILGLUTHRANTYRTHARGILEPHELEUTYRSEGLYGLUPROTH 226
DB 482 TTCAAGAACATGAAACCACTATACAGAAATTTCTTTACAGTGAACCTATTA 541
QY 227 LEUGLYASNGIUTHSERVALPHEGLYPROTHRGYASNVLEUTHLEUGLYLEUA 246
DB 542 CTGGGAAATGAAACATCTGTTTGGGCAAGGAAACAAAGCTCTGTTTGAAGCA 601
QY 247 LYRARGPHELYRTPROPHELYSPROHISLEU-PROTHRLYSGIUPHELEUSE 266
DB 602 AAAAGATTTTATTAACCTTCAAGCAATTTGGCCACATTAAGAAATTT--CT 658
QY 266 ULEUGIUNILEPHEASPALVALILEVALHISLYSGINPHELYRLEUPHELY 286
DB 659 CTCTGCAATTTTGAATGCTGATTTGCCCCCAGCCGTCCTATTTCT--TATA 717
QY 286 UTYRTPHELEUPROHELYSPHE 294
DB 718 -TATGGGTTTACCTATGAAATTC 741

```

RESULT 10  
 LOCUS B1090566 869 bp mRNA linear EST 20-JUN-2001  
 DEFINITION 60285673F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:4996891 5',  
 mRNA sequence.  
 ACCESSION B1090566  
 VERSION B1090566  
 KEYWORDS EST, B1090566.1 GI:14508896  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE Homo sapiens  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 TITLE 1 (bases 1 to 869)  
 JOURNAL NIH-MGC http://mgs.nci.nih.gov/  
 COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgsaps-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC Clone distribution information can be  
 found through the I.M.A.G.E. Consortium/MLNL at:  
 http://image.llnl.gov  
 Plate: L1AM1023 row: a column: 20  
 High quality sequence stop: 843.  
 Location/Qualifiers  
 1..869  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4996891"  
 /cell\_line="MGC36"  
 /lab\_host="DH10B"  
 /clone\_id="NIH\_MGC\_10"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo 3T.  
 Average insert size 1.5 kb. Library prepared by Life  
 Technologies."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 4.87e-122 Length: 869  
 Score: 1191.00 Matches: 232  
 Percent Similarity: 93.93% Conservative: 0  
 Best Local Similarity: 93.93% Mismatches: 8  
 Query Match: 67.75% Indels: 8  
 DB: 12 Gaps: 0  
 US-10-010-050a-2\_COPY\_31\_346 (1-316) x B1090566 (1-869)

1 ArgValSerGlyIleProSerArgArgHisTTPProValProTyrLysArgPheAspPhe 20  
 142 CGGGTCTCGGGGATCCCTCCCGGCGCAGCTGGCGTGCCTACAGCGCTTACATTC 201  
 21 ArgProIySProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40  
 202 CCTCCAAACCTGATCCTTATGTCAGAGTAAGTAACTTCTGTCCACATCGGCTCACCT 261  
 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTTP 60  
 262 ATCCAGATTATGAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 321  
 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysGlyIleMetHisAspAlaIleGlyPhe 80  
 322 GAATTAAATATGAGAGACCTCCCTGGGACACTGAAATATATGATGATGATGATGATG 381  
 81 ArgSerThrLeuThrGlyLysLeuTyrThrMetGluTTPTrpGlyLeuPheGlnLeuGly 100  
 382 AGAAGTAACTTAACTGCGCAAGAACTTCAAAATGAGATGATGATGATGATGATGATG 441

QY 101 AsnGlyThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120  
 DB 442 AACTGTAACTTCCCACTCCGACCTGAATGATGATGATGATGATGATGATGATGATGATG 501  
 QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTTPProTyrLysArgThrLeuVal 140  
 DB 502 GCTGCTCTCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 561  
 QY 141 GlnValAlaThrTleSerGlyAsnMetPheAsnGlnMetAlaLysTyrValLysGlnAsp 160  
 DB 562 CAAATGACAACTAATTCAGAAACATGTTCAACCAATGCGC-AGGTGGGTGAAACAGGAC 620  
 QY 161 AsnGluThrGlyIleTyrTyrGluThrPheAsnValLysAlaSerProGluLysGlyAla 180  
 DB 621 AATGAAACGGAATCTATATGACATGAAATGAAAGCCAGCCACAGAAACAGCGCGC 680  
 QY 181 --GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysL 200  
 DB 681 AGGAGACATGCTTATGATTCCTACGACTGTCCAAATTTGTGTTAAGACCTTTAACAGT 740  
 QY 200 euAlaGlu-PheGlyAlaGluPheLysAsnIle-GluThrAsnTyrThrArgIlePheLe 219  
 DB 741 TGGCTGAACCTGGGAGCAGAGTTCAGAAACATTAGAAACCAANTATACAAAGATATTTCT 800  
 QY 219 uTyrSerGly-GluProThrTyrLeuGly-AsnGluThr-SerValPheGlyProThrGly 238  
 DB 801 TTACAGTGAAGAACTACTTATCTGGAATGAAACCATCTCTTTAGGCGGCAACCAAG 860  
 QY 238 YAsnLys 240  
 DB 861 AATCAAG 867

RESULT 11  
 LOCUS BX370641 908 bp mRNA linear EST 08-MAY-2003  
 DEFINITION BX370641 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CS0DL004YP03 5-PRIME, mRNA sequence.  
 ACCESSION BX370641 GI:30459737  
 VERSION BX370641.1  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 7238.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BAG053ZH03 CS05016.1&cluster=7238.f.  
 Contact : Feng Liang Email : fliang@life.techn.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0BAG053ZH03\_CS05016\_1.  
 Location/Qualifiers  
 1..908  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL004YP03"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and Ecor V

sites of the pcwvSPORT 6 vector. Library was normalized."

ALIGNMENT Scores: 2.44e-119 Length: 908  
 Tred. No.: 1167.00 Matches: 217  
 Percent Similarity: 96.88% Conservative: 0  
 Percent Similarity: 96.88% Mismatches: 7  
 Percent Match: 66.38% Indels: 1  
 Gaps: 0

JS-10-010-050A-2\_COPY\_31\_346 (1-316) x BX370641 (1-908)

2Y TTPYTGULUePheGlnLeuGlyAsnGlyThrPheProHisLeuArgProGluMetAsp 1-2  
 3B TGGCTGACCTTTCCTCACTGGCAACTGTACATTTCCCACTCCGACCTGCA-ATGAT 62  
 4 TGGCTGACCTTTCCTCACTGGCAACTGTACATTTCCCACTCCGACCTGCA-ATGAT 62  
 2Y 113 A1aProPheTTPCyAsnGlnGlyAlaAlaCysPhePheGlnGly1LeaAspValHis 132  
 3B 63 GCCCTTTCCTGCTGTATCAAGGGCTGCTGCTTTTGGAGGCAATTGATGATGTTTCA 122  
 2Y 133 TTPYSGLUAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 152  
 3B 123 TCGAAGCAAAATGGGACATTAAGTTCAAGTACACTATATCGAAGACATGTTCAACCA 132  
 2Y 153 MetAlaLysTTPValLysGlnAspAsnGluThrGlyIleTyrTyrGluThrTTPAsnVal 172  
 3B 183 ATGGCAAAAGTGGGTAAACAGACATTAAGACAAATTTATTAAGACACATGATGTA 242  
 2Y 173 LysAlaSerProGluLysGlyValGluThrTTPPheAspSerTyrAspCysSerLysPhe 192  
 3B 243 AAGGCAAGCCGCAAAAGGGGCAAGACATGATGTTGATTCCTACAGCTTCCTCAATTT 302  
 2Y 193 ValLeuArgThrPheAsnLysLeuAlaGluLysPheGlyAlaGluPheLysAsnIleGluThr 212  
 3B 303 GTGTTAAGACCTTTAAACAGTTGCTGATTTGGAGCAAGTTTCAAGAACTAAGAAAC 362  
 2Y 213 AsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSer 232  
 3B 363 AACTATACAAAGATATTTCTTTTACAGTGAAGAACCTACTATCTGGAAATGAACATCT 422  
 2Y 233 ValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrPro 252  
 3B 423 GTTTTGGGCAACAGAAAGAACTCTGTTTGGCTTAAAGATTTATTAATATTAATATTA 482  
 2Y 253 PheLysProHisLeuProThrLysGluLysPheLeuSerLeuLeuGlnIlePheAspAla 272  
 3B 483 TTCAAACCACTTTGGCAACTAAAGAAATTTCTGTGAGTCTTTCGAAATTTTGAATGA 542  
 2Y 273 ValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTTPPheLeuPromet 292  
 3B 543 GTATATGTGCAAAACAGTTCTTCTTTTATTAATTTGAAATTTGTTTAACTAAG 602  
 2Y 293 LysPheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsnLysThr 312  
 3B 603 AATTTCCCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 662  
 2Y 313 LeuSerGlyLeu 316  
 3B 663 CTCTCTGCTTTA 674

RESULT 12 974 bp MRNA linear EST 20-AUG-2002  
 LOCUS B0919385 A8802804 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6382620  
 DEFINITION 5', mRNA sequence.  
 ACCESSION B0919385  
 VERSION B0919385.1 GI:22334083  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 974)  
 AUTHORS NIH-MGC  
 TITLE NIH-MGC  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: DCTD/DRP  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LNCM2574 row: 1 column: 13  
 High quality sequence stop: 634.  
 Location/Qualifiers  
 1..974

#### FEATURES

source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6382620"  
 /tissue\_type="carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 40"  
 /note="Organ: prostate; Vector: pOT7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

#### ORIGIN

ALIGNMENT Scores:

Pred. No.: 2.08e-118 Length: 974  
 Score: 1159.00 Matches: 213  
 Percent Similarity: 83.28% Conservative: 26  
 Best Local Similarity: 74.22% Mismatches: 35  
 Query Match: 65.93% Indels: 13  
 DB: 13 Gaps: 3

US-10-010-050A-2\_COPY\_31\_346 (1-316) x B0919385 (1-974)

QY 7 SerArgArgHisTTPProValProTyrLysArgPheAspPheArgProLysProAspPro 26  
 DB 117 TCTGGGCAACGCTGGCGGGTCCCTTCAAGCCCTTCTTCCGTCGGAAGACAGATCCC 176  
 QY 27 TTPYSGLUAlaLysTyrThrPheCysProThrGlySerProIleProValMetGluGly 46  
 DB 177 TACTGTCAAGCTTAAGATATTTCTGTCTTACCGGCTGCCCATCCCACTTATGAAGAC 236  
 QY 47 AspAspAspIleGluValPheArgLeuGlnAlaProValTTPGluPheLysTyrGlyAsp 66  
 DB 237 AATGACGTATGAGGCTTACGATCAACGCCCATTTGGGAATTTAAATATGAGAC 296  
 QY 67 LeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGly 86  
 DB 297 CTCCTGGACACTTTTAACTTATGATGACGCGGTGGATTCAGGACACACTGACAGGC 356  
 QY 87 LysAsnTyrThrMetGluTTPYrGluLeuPheGlnLeuGlyAsnGlyThrPheProHis 106  
 DB 357 AAGAACTACCAATTTAAGTGTATGAACTTTCCACTGGGCAACGTACATTCCCCAC 416  
 QY 107 LeuArgProGluMetAspAlaProPheTTPCyAsnGlnGlyAlaAlaCysPhePheGlu 126  
 DB 417 CTCGGGCTGCAAGAGCGCTCCCTCTGCTGTAAACCAAGGGGCAAGCTGCTTTTGA 476  
 QY 127 GlyTLeAspAspValHisTTPYrGluAsnGlyThrLeuValGlnValAlaThrIleSer 146  
 DB 477 GGAATAGATGATTAACACCTGGAAGAAACGGGACACTGTGCTTGGCAACATATCC 536  
 QY 147 GlyAsnMetPheAsnGlnMetAlaLysTTPValLysGlnAspAsnGluThrGlyIleTyr 166

```

Db      537  GGAAGCATTATACAAAGTGGCCGAGTGGTGAACAGACAAATGAACTGGGATTAT 596
Qy      167  TTTGTTTThrTTPaenVallysalaserProGluLysGlyAlaGluThrTTPheaspSer 186
Db      597  TATGAGCAATGAGACGGTCCAGCCGAGCAGCAAGGGGCCAGACAGTGGTTCGAGTCC 656
Qy      187  TTTAspCysSerlyspheValleuArgThrPheasnlyleuAlaGluPheGlyAlaGlu 206
Db      657  TAGCACTGTTGCAATTTTGTCTTAAAGACATATAGAAATTTGCTGATTTGGACAAGAA 716
Qy      207  PheylsAnlllegluthrAsnlyrThrArgilepheleutySerGlyGluProThrTyr 226
Db      717  TTCAAGAGATAGAAACAACTATACGAAATATTTCTTACAGTGAAGAGCTATTATAC 776
Qy      227  LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnlyserThrLeuGlyLeuAla 246
Db      777  CTGGGAATAGAAACATCTATTTTGGCCCAAGAGAAACAGCTTGGCTTGGCCAT 836
Qy      246  e-LysArgPheTyrTyrProPheLyserProHisleu---ProThrLyseGlu-Pheleu 264
Db      837  AAAAAAATTTTATGTCCTCCCTCCAGACCGAATTTGTCAACCAAGATTTTCTGATG 896
Qy      265  SerleuLeuGln---llepheaspAlaVal-----lleVal 275
Db      897  AAATTTCTTGAATAATTTTGTATACGTGATATATACCCCGAGGTTCTTACTTGT 356
Qy      276  HislyseGlnPheTyr 280
Db      957  TTTAAACCTTTGG 971

RESULT 13
LOCUS      BI080303                      915 bp      mRNA      linear      EST 20-JUN-2001
DEFINITION 602876736f1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5008544 5',
            mRNA sequence.
ACCESSION  BI080303
VERSION    BI080303.1  GI:14498633
KEYWORDS   EST.
SOURCE      Mus musculus (house mouse)
ORGANISM   Mus musculus
REFERENCE   Mus musculus
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
COMMENT    NIH-MGC http://mgi.nci.nih.gov/.
            1 (bases 1 to 915)
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LLM11053 row: 9 column: 09
            High quality sequence strop: 748.
            Location/Qualifiers
                1..915
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="FVB/N-3"
                /db_xref="taxon:10090"
                /clone="IMAGE:5008544"
                /tissue_type="tumor, biopsy sample"
                /dev_stage="5 months"
                /lab_host="DH10B"
                /clone_1ib="NCI CGAP Mam2"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigator:
                providing samples: Gilbert Smith, NIH"

```

```

ORIGIN
Alignment Scores:
Pred. No.:      2,48e-118
Score:          1158.00
Percent Similarity: 84.83%
Best Local Similarity: 75.86%
Query Match:    65.87%
DB:             12
Gaps:           1

US-10-010-050A-2_COPY_31_346 (1-316) x BI080303 (1-915)
Qy      25  AAPPPTTTCyGlnAlaValTyrThrPheCysProThrGlySerProIleProValMet 44
Db      25  GATCCTACTGTCACAGTAAGTACTTCTTCTTACCGGCTCGCCATCCAGTTATG 84
Qy      45  GluGlyAspAspAsp11legluValPheArgLeuGlnAlaProValTTPGluPheLy 64
Db      85  AAGGACAAATGAGCTCATCGAGGTCTTACGACATCAAGCCCGATTTGGGAATTAATAT 144
Qy      65  GlyAspLeuLeuGlyHisleuLysIleMetHisAspAlaAlleglyPheArgSerThr 84
Db      145  GGAGACCTCCTCGGACACTTAACTTATGCAATGACGAGCGGTGGAAATTCAGAGCACTG 204
Qy      85  ThrGlyLyAsnThrThrMetGluTTPYrGluLeuPheGluLeuGlyAsnCyThrPhe 104
Db      205  ACAGGCAAGAACTACACAAATGATGATGATGATGATGATGATGATGATGATGATGAT 264
Qy      105  ProHisLeuArgProGluMetAspAlaProPheTTPCyAsnGlnGlyAlaAlaCyPhe 124
Db      265  CCCCACTCCTCGGCTGCAAGAGCGCTCCCTTGGGTGATCAAGAGGCGAGCTGCTTT 324
Qy      125  PheGluGlyTleAspAspValHisTTPYrGluAsnGlyThrleuValGluValAlaThr 144
Db      325  TTTGAAAGATATAGATGATTAACCTGAGAAAGAGGACACTGTCATGCTTGCACC 384
Qy      145  IleSerGlyAsnMetPheAsnGlnMetAlaLyserTTPYrVallyseGlnAspAsnGlu 164
Db      385  ATATCCGAAACACATTTAACAAAGTGGCCGAGTGGTGAAGACGAAATGAACTGGGG 444
Qy      165  IleTyrTyrGluThrTTPaenVallysalaserProGluLysGlyAlaGluThrTTPhe 184
Db      445  ATTATTTATAGACATGACGAGCGTCCAGCCGAGCAAGAGGCGCAGACCTGCTTC 504
Qy      185  AspSerTyrAspCysSerlyspheValleuArgThrPheasnlyleuAlaGluPheGly 204
Db      505  GAGTCTTACGACCTGTTGAAATTTGTCTTAAAGCATATAGAAATCGCTGAATTTGGA 564
Qy      205  AlaGluPheLyserAnlllegluthrAsnlyrThrArgilepheleutySerGlyGluPro 224
Db      565  ACAGAAATTCAGAAAGATAGAAACAACTATACGAAATATTTCTTACAGTGAAGGCT 624
Qy      225  ThrTyrleuGlyAsnGluThrSerValPheGlyProThrGlyAsnlyserThrleuGly 244
Db      625  ATTATACCTGGGAATTAAGAACTATTTTGGCCCAAGGAAACAGACTGCTGCTTGG 684
Qy      245  AlaIleLyserArgPheTyrTyrProPheLyserProHisleuProThrlyGluPheLeu 264
Db      685  GCCATAAAAAAACCTTATAGCCCTTCAGACCGTA-TTGTCAACCAAGAGTTCTGATG 743
Qy      265  SerleuLeuGlnllepheaspAlaValleuValHis-LyseGlnPheTyrleuPheTyrAs 284
Db      744  AATTC-TTGACAACTTTGATACAGTATTTTACCGCAAGAGGTCTACTGTTTATATA 802
Qy      284  nPheGluTyrTTPheleuProMetLysePheProPheIleLyserThrTyrTTPGluGlu 304
Db      803  CTTGAGATA-TGGGTTCACATGAAA---CCCTTGTCAAAATTAATACAGAGAAAC 858
Qy      304  eProleuProIleArgAsnlyrThrleu 313
Db      859  -CCGGTTACGATCGAAATTCAGGTTA 865

RESULT 14

```







## Alignment Scores:

Pred. No.:	4.52e-116	Length:	641
Score:	1136.00	Matches:	200
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	64.62%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050a-2\_copy\_31\_346 (1-316) x CB215456 (1-641)

```

QY      1  ArgValSerGlyIleProSerArgArgHisTrpProValProTyrIlySAArgPheASPhe 20
DB      42  CGGGCTCGGGCATCCCTCCCGGGCCACCTGGCCCGGCTTCAAGCGCTTGACTTC 101
QY      21  ArgProIlySPROASPProTyrCysGlnAlaIlySTyrPheCysProThrGlySerPro 40
DB      102  CGTCCAAACCTGATCTTATTTGTCAGCTAGTACTTCTGTCTCAACTGGCTCACT 161
QY      41  IleProValMetGluGlyASPASPaspIleGluValPheArgLeuGlnAlaProValTrp 50
DB      162  ATCCCAAGTTATGGAGGGTGAATGATGATGACATTGAGTTTTTGATTACAAGCCCGATGG 221
QY      61  GluPheIlySTyrGlyASPLeuLeuGlyHisLeuIlySileuEHisASPAlaIleGlyPhe 30
DB      222  GAATTTAAATGAGACCTCTCTGGACACTTGAAATTTATGCATGATGCCATTGGATTTC 281
QY      81  ArgSerThrLeuThrGlyIlyASnTyrThmerGluTrpTyrGluLeuPheGlnLeuGly 100
DB      282  AGAAGTACATTACTGGCAGACACTACACAAATGGAATGGATGATGACTTTTCCAACTTGGC 341
QY      101  AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
DB      342  AACTGTACATTTCCTCCCATCTCCGACCTGAATGATGATGCCCTTCTGTGTATCAAGGC 401
QY      121  AlaAlaCysPhePheGluGlyIleASPaspValHisTrpIlySGluAsnGlyThrLeuVal 140
DB      402  GCTGCCCTGCTTTTGGAGGAAATGATGATGCTTCACCTGGAAGGAAATGGGACATTAGTT 461
QY      141  GlnValAlaThrIleSerGlyASnMetPheAsnGlnMetAlaIlySTrpValIlySGlnASP 160
DB      462  CAAAGTACACTATATACGAAACATGTTCAACCAAAATGGCAATGGGTGAACAGGAC 521
QY      161  AsnGluThrGlyIleIlyTyrGluThrTrpAsnValIlySalaserProGluIlySGlyAla 180
DB      522  AATGAAACAGGAATTTATATGAGACATGGAATGTAAAGCCAGCCCAAGAAAAGGGGCA 581
QY      181  GluThrTrpPheASPserTyrASPcysSerIlySPheValLeuArgThrPheAsnIlySleu 200
DB      582  GAGCATGGTTGATTCCTAAGACTGCTCAAAATTTGTGTTAAGAGACCTTTAAACAACTTG 641

```

Search completed: April 25, 2004, 05:46:13

Job time : 4028.01 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:04:07 ; Search time 629.806 Seconds  
(without alignments)  
2262.168 Million cell updates/sec

Title: US-10-010-050A-2\_COPY\_31\_346  
Perfect score: 1/58  
Sequence: 1 RVSGIPSRHMPVPRKPDF.....IKITYEIPIDINKTUSGL 316

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2907579 seqs, 225431464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1h  
-Q=/cgn2\_1/USPTO.spool/US10010050/runat\_22042004\_113205\_27584/app\_query.fasta\_1.1372  
-DB=published Applications NA -QMT=fastap -SUFFIX=rmpb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bits -STRAT=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.ccd -LIST=45 -DOCLIGN=200 -THR SCORE=pcp -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcp -NORM=ext -HEADSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USR=US10010050 @CNG 1 1 776 @runat\_22042004\_113205\_27584  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DEBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:\*

- 1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/prodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/prodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 17: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 19: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	1758	100.0	1486	9	US-09-122-383-1	Sequence 1, Appl1
2	1758	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appl1
3	1747	99.4	1751	13	US-10-653-595-26	Sequence 26, Appl1
4	1747	99.4	1751	13	US-09-397-945-26	Sequence 26, Appl1
5	1515	86.2	1038	9	US-09-122-383-13	Sequence 13, Appl1
6	1515	86.2	1038	14	US-10-010-050A-13	Sequence 13, Appl1
7	886	50.4	494	9	US-09-864-761-22592	Sequence 825, App
8	727	38.9	506	14	US-10-040-739-825	Sequence 5832, App
9	684.5	38.9	474	9	US-09-864-761-5832	Sequence 1859, App
10	472	26.8	497	16	US-10-264-049-1859	Sequence 2579, App
11	359.5	20.4	351	9	US-09-867-701-2579	Sequence 1363, App
12	241	13.7	473	9	US-09-954-456-1363	Sequence 10339, A
13	121	6.9	60	10	US-10-198-846-9753	Sequence 36401, A
14	104.5	5.9	2667	15	US-10-359-493-36401	Sequence 21, Appl
15	97.5	5.5	986	16	US-09-822-846-21	Sequence 159, App
16	96	5.5	3580	10	US-09-374-046A-159	Sequence 127, App
17	96	5.5	3580	13	US-10-616-263-159	Sequence 127, App
18	96	5.5	14070	9	US-09-870-759-127	Sequence 127, App
19	94.5	5.4	14070	10	US-09-751-708A-127	Sequence 22565, A
20	94.5	5.4	3717	13	US-10-282-122A-22565	Sequence 9, Appl1
21	94	5.3	1612	14	US-10-003-392-9	Sequence 19, Appl1
22	92.5	5.3	2115	14	US-10-003-392-19	Sequence 120013, A
23	92.5	5.2	1468	13	US-10-424-599-120013	Sequence 3, Appl1
24	92	5.2	7680	10	US-09-953-318-3	Sequence 214, App
25	91.5	5.2	7680	13	US-10-211-462-214	Sequence 711, App
26	91.5	5.2	7680	13	US-10-342-687-711	Sequence 711, App
27	91.5	5.2	7680	13	US-10-172-118-711	Sequence 16, Appl1
28	91.5	5.2	7680	13	US-10-446-373-3	Sequence 77, Appl1
29	91.5	5.2	7680	15	US-10-021-660-16	Sequence 77, Appl1
30	91.5	5.2	7680	15	US-10-393-590-77	Sequence 109423, A
31	91.5	5.2	7680	15	US-10-393-567-77	Sequence 26859, A
32	91.5	5.2	7680	15	US-10-394-087-77	Sequence 1791, App
33	89	5.1	2131	13	US-10-424-599-109423	Sequence 253, App
34	88.5	5.0	870	16	US-10-369-493-26859	Sequence 4, Appl1
35	88.5	5.0	3035	15	US-10-108-260A-1791	Sequence 253, App
36	88.5	5.0	5574	16	US-10-037-270-253	Sequence 23914, A
37	88.5	5.0	5574	16	US-10-117-722-253	Sequence 22862, A
38	88.5	5.0	1317	15	US-10-028-245-4	Sequence 43789, A
39	88	5.0	1815	13	US-10-282-122A-23914	Sequence 16281, A
40	88	5.0	1826	15	US-10-028-245-1	Sequence 36547, A
41	88	5.0	9585	13	US-10-282-122A-22862	
42	87	4.9	1167	16	US-10-369-493-43789	
43	87	4.9	1314	13	US-10-282-122A-16281	
44	87	4.9	3488	16	US-10-369-493-36547	
45	87	4.9				

## ALIGNMENTS

RESULT 1  
US-09-122-383-1  
Sequence 1, Application US/09122383A  
Patent No. US20020042093A1  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Gilbertson, Debra G.  
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
FILE REFERENCE: CHROMOSOME 13  
CURRENT APPLICATION NUMBER: US/09/122,383A  
CURRENT FILING DATE: 1998-07-24  
EARLIER APPLICATION NUMBER: 60/053,613  
EARLIER FILING DATE: 1997-07-24  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1486  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (47)...(1084)

US-09-122-383-1

## Alignment Scores:

Pred. No.:	2,28e-222	Length:	1486
Score:	1758.00	Matches:	316
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050A-2\_COPY\_31\_346 (1-316) x US-09-122-383-1 (1-1486)

```

QY 1 ArgValSerGlyIleProSerArgHsiTrpProValProTyIysArgPheasphe 20
DB 137 CGGGTCGCGGACATCCCTCCCGGCCCACTGGCCGGTCCCTCAAGAGCCTTGACTTC 196
QY 21 ArgProIysProAspProTyIysGlnAlaIysTyIThrPheCybProThrgIysPro 40
DB 197 CGTCCAAAACCTGATCTTATGTCAAGCTAAGTATACCTTCTGTCCAACTGGCTCACC 256
QY 41 IleProValMetGlnGlyAspAspApIleGlnValPheArgGlnAlaProValTrp 60
DB 257 ATCCCAAGTATGAGGGGTGATGATGACATTGAAGTTTTCGATTACAAAGCCCAATG 316
QY 61 GluPheIysTyIAspLeuLeuGlyHsiLeuIysIleMetHsiAspAlaIleGlyPhe 60
DB 317 GAATTTAAATATGAGACCTCTCGGACACTGAAATATATGATGATGATGATGATG 376
QY 81 ArgSerThrLeuThrgIyIysAsnTyIThrMetGluTrpTyIleuPheGlnLeuGly 100
DB 377 AGAAGTACATTAATCTGGCAAGAACTACACAAATGAAATGATGATGATGATGATG 436
QY 101 AsnCySerThrPheProHsiLeuArgProGlnMetAspAlaProPheTrpCybAsnGlnGly 120
DB 437 AACTGTCATTTCCCGCATCTCCGACCTGGAATGATGATGATGATGATGATGATG 496
QY 121 AlaAlaCySPhePheGlnGlyIleAspAspValHsiTrpIysGlnGlnGlyThrLeuVal 140
DB 497 GCTCCCTGCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATG 556
QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIysTrpValIysGlnAsp 160
DB 557 CAAGTACCACTATATCAGGAACATGTCAACCAATGCAAGTGGGAAACAGAGAC 616
QY 161 AsnGluThrgIyIleTyIThrGluThrTrpAsnValIysAlaSerProGlnIysGlyAla 180
DB 617 AATGAAACAGGAATTTATGAGAACATGATGATGATGATGATGATGATGATGATG 676
QY 181 GluThrTrpPheAspSerTyIAspCySerIysPheValIleuArgThrPheAsnIysLeu 200
DB 677 GAGACATGTTGATCTCTAGCATGCTTCCAAATTTGTGTTAAGACCTTTAACAAGCTTG 736
QY 201 AlaGluPheGlyAlaGluPheIysAsnIleGluThrAsnTyIThrAsnGlnIlePheLeuTyI 220
DB 737 GCTGAATTTGAGCAGAGTTCAAGAACATGAAACCAACTATACAAAGATATTTCTTTAC 796
QY 221 SerGlyGluProThrTyIleuGlyAsnGluThrSerValPheGlyProThrGlyAsnIys 240
DB 797 AGTGAAGACCTACTTACTCTGGAATGAAACATCTGTTTGGCCCAACGAGAAACAG 856
QY 241 ThrLeuGlyLeuAlaIleIysArgPheTyIThrProPheIysProHsiLeuProThriIys 260
DB 857 ACTCTTGTTTGGCCTTAAAGATTTTATACCCCTTCAAAACCAATTTCCCAACTAA 916
QY 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHsiIysGlnPheTyI 280
DB 917 GAATTTCTGTTGAGCTCTTGCATTTTGTGATGAGATGATGATGATGATGATGATG 976
QY 281 LeuPheTyIAsnPheGluTyIThrPheLeuProMetIysPheProPheIleIysIleThr 300
DB 977 TTGTTTAAATTTGAAATATGATGATGATGATGATGATGATGATGATGATGATG 1036
QY 301 TyrGluGluIleProLeuProIleArgAsnIysThrLeuSerGlyLeu 316

```

DB 1037 TATGAGAAATCCCTTACTATCAGAAACAAACACTCTCTGTTTA 1084

## RESULT 2

US-10-010-050A-1

```

; Sequence 1, Application US/10010050A
; Publication No. US20020173624A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; TITLE OF INVENTION: CHROMOSOME 13
; FILE REFERENCE: 97-38C1
; CURRENT APPLICATION NUMBER: US/10/010,050A
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 09/122,383
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/053,613
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (47)...(1084)
US-10-010-050A-1

```

## Alignment Scores:

Pred. No.:	2,28e-222	Length:	1486
Score:	1758.00	Matches:	316
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050A-2\_COPY\_31\_346 (1-316) x US-10-010-050A-1 (1-1486)

```

QY 1 ArgValSerGlyIleProSerArgHsiTrpProValProTyIysArgPheasphe 20
DB 137 CGGGTCGCGGACATCCCTCCCGGCCCACTGGCCGGTCCCTCAAGAGCCTTGACTTC 196
QY 21 ArgProIysProAspProTyIysGlnAlaIysTyIThrPheCybProThrgIysPro 40
DB 197 CGTCCAAAACCTGATCTTATGTCAAGCTAAGTATACCTTCTGTCCAACTGGCTCACC 256
QY 41 IleProValMetGlnGlyAspAspApIleGlnValPheArgGlnAlaProValTrp 60
DB 257 ATCCCAAGTATGAGGGGTGATGATGACATTGAAGTTTTCGATTACAAAGCCCAATG 316
QY 61 GluPheIysTyIAspLeuLeuGlyHsiLeuIysIleMetHsiAspAlaIleGlyPhe 80
DB 317 GAATTTAAATATGAGACCTCTCGGACACTGAAATATATGATGATGATGATGATG 376
QY 81 ArgSerThrLeuThrgIyIysAsnTyIThrMetGluTrpTyIleuPheGlnLeuGly 100
DB 377 AGAAGTACATTAATCTGGCAAGAACTACACAAATGAAATGATGATGATGATGATG 436
QY 101 AsnCySerThrPheProHsiLeuArgProGlnMetAspAlaProPheTrpCybAsnGlnGly 120
DB 437 AACTGTCATTTCCCGCATCTCCGACCTGGAATGATGATGATGATGATGATGATG 496
QY 121 AlaAlaCySPhePheGlnGlyIleAspAspValHsiTrpIysGlnGlnGlyThrLeuVal 140
DB 497 GCTCCCTGCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATG 556
QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIysTrpValIysGlnAsp 160
DB 557 CAAGTACCACTATATCAGGAACATGTCAACCAATGCAAGTGGGAAACAGAGAC 616
QY 161 AsnGluThrgIyIleTyIThrGluThrTrpAsnValIysAlaSerProGlnIysGlyAla 180

```

617 AATGAAACAGGAATTTATATGACATGAAATGTAAGCCACCAAGAAAGGGGGCA 676  
181 GlnThrTrpPheAspSerTyrAspCysSerIlePheValLeuArgThrPheAsnIleLeu 260  
677 GAGCAATGGTTGATTCCTACGACTGTTCCAAATTTGTGTTAGACCTTTAAACAATTG 756  
201 AlaGluPheGlyAlaGluPheGlyAsnIleGluThrAsnIleThrArgIlePheLeuTyr 220  
737 GCTGAATTTGGAGCAGATTCAAGACATGAAACCAATATACATAAATAATTTCTTTAC 756  
221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIle 240  
797 AGTGAAGAACTCTTATCTGGAATGAAACATCTGTTTGGCCCAACGGAACAG 856  
241 ThrLeuGlyLeuAlaIleValArgPheTyrTyrProPheIleProHisIleProThrIle 260  
857 ACTCTGGTTTACCCATTAAGAAATTTTATTAACCCCTTCAACCAATTTGCCAATCAA 956  
261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValIleValGlnPheTyr 280  
917 GAATTTCTGTTGAGCTCTTGCAAAATTTTGATGAGATGATGCAACAACAGTTCTAT 976  
281 LeuPheTyrAsnProGluTyrTyrPheLeuProMetIlePheProPheIleValIleThr 300  
977 TTGTTTATTAATTTTGAATATGTTTATTTACCTATGAATATCCCTTTATTAATAATAACA 1036  
301 TyrGluGluIleProLeuProIleArgAsnIleThrLeuSerGlyLeu 316  
1037 TATGAGAAATCCCTTATCTATCGAATAAACAACATCTCTGTTTA 1084

## RESULT 3

US-10-653-595-26

Sequence 26, Application US/10653595

Publication No. US20040048304A1

GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

FILE REFERENCE: P2027P1

CURRENT APPLICATION NUMBER: US/10/653,595

CURRENT FILING DATE: 2003-09-03

PRIOR APPLICATION NUMBER: US 09/397945

PRIOR FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: 60/078,566

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,576

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,573

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,574

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,579

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/080,314

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/080,312

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 470

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 26

LENGTH: 1751

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (1520)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE  
LOCATION: (1557)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1689)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1729)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1735)  
OTHER INFORMATION: n equals a,t,g, or c  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1741)  
OTHER INFORMATION: n equals a,t,g, or c  
US-10-653-595-26

## Alignment Scores:

Pred. No.: 8,57e-221 Length: 1751  
Score: 1747.00 Matches: 314  
Percent Similarity: 99.37% Conservative: 0  
Best Local Similarity: 99.37% Mismatches: 2  
Query Match: 99.37% Indels: 0  
DB: Gaps: 0

US-10-010-050A-2\_COPY\_31\_346 (1-316) x US-10-653-595-26 (1-1751)

1 ArgValSerGlyIleProSerArgArgHisThrProValProTyrIleArgPheAspPhe 20  
139 CGGGTCTGGGCAATCCCTCCCGGGCCACGCGCGGCGCTTACAGCGCTTACATTC 198  
21 ArgProIleProAspProTyrCysGlnAlaIleTyrThrPheCysProThrIleSerPro 40  
199 GCTCCAAACCTGATCCTTATTTGCAAGTATATCTTCTGTCGCAACTGGCTCACCT 258  
41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60  
259 ATCCAGTTATGAGGGGTGATGATGACATTAAGTTTTCATTTCAGGCTTCAAGCCCAATATG 318  
61 GluPheIleTyrGlyAspLeuLeuGlyHisLeuValIleMetHisAspAlaIleGlyPhe 80  
319 GAATTTAATATGAGGACCTCTGGGACACTTGAATAATATGACATGATCCATTGATTC 378  
81 ArgSerThrLeuThrGlyIleAsnIleTyrMetGluTyrTyrGluLeuPheGlnLeuGly 100  
379 AGAATGATCATTTACCTGCGAAGAACTACCAATGAAATGATGATGAACTTTCCAACTTGGC 438  
101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120  
439 AACTGTACATTTCCCAATCTCCGACCTGMAAAGATGCCCTTCTGCTGATATCAAGGC 498  
121 AlaAlaCysPhePheGluGlyIleAspAspValHisTyrPheGluValAsnIleThrLeuVal 140  
499 GCTGCTGCTTTTGGAGGAATTTGATGATGATCTTCACTGGAAGAAATGCGACTTAACT 558  
141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValIleGlnAsp 160  
559 CAAAGTACCACTTATACGAAACATGTTCAACCAATGCAAAAGGCGGTGAACAGGAC 618  
161 AsnGluThrGlyIleTyrTyrGluThrTyrAsnValIleAlaSerProGluIleGlyAla 180  
619 AATGAAACAGGAATTTATATGACATGAAATGTAAGAAACCAAGCCCAAGAAAGGGGGCA 678  
181 GluThrTrpPheAspSerTyrAspCysSerIlePheValIleuArgThrPheAsnIleLeu 200  
679 GAGACATGGTTGATTCCTACGACTGTTCCAAATTTGTGTTAAGACCTTTAAACAAGTTG 738  
201 AlaGluPheGlyAlaGluPheGlyAsnIleGluThrAsnIleTyrThrArgIlePheLeuTyr 220

Db 739 GGTGAATTGGAGCAGAGTTCAAGACATGAAACCACTATACAGATATTTCTTAC 798  
Qy 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240  
Db 799 AGTGGAGAACCTACTATCTGGAATGAAACATCGTTTTGGGCCAACAGAAACAAG 858  
Qy 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260  
Db 859 ACTCTTGTTTGGCCATTAAGATTTTATTTACCCCTTAACACCATTTGGCAACTAA 918  
Qy 261 GluPheLeuLeuSerLeuLeuGluIlePheAspAlaValIleValHisLysGluPheTyr 280  
Db 919 GAATTCCTGTGAGTCTCTTGCAATTTTGATGAGAGATGTGCACAAACAGTTCTAT 978  
Qy 281 LeuPheTyrAsnPheGluTyrTyrPheLeuProMetLysPheProPheIleLysIleThr 300  
Db 979 TTGTTTATATATTTTGAATATGTTTATCCATGTAATTCCTTTATTAATAATACA 1038  
Qy 301 TyrGluGluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 316  
Db 1039 TATGAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGGTTTA 1086

RESULT 4  
US-09-397-945-26  
; Sequence 26, Application US/09397945  
; Publication No. US20030065139A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc. et al.  
; TITLE OF INVENTION: 95 Human secreted proteins  
; FILE REFERENCE: P2027P1  
; CURRENT APPLICATION NUMBER: US/09/397,945  
; CURRENT FILING DATE: 1999-09-17  
; PRIOR APPLICATION NUMBER: PCT/US99/05804  
; PRIOR FILING DATE: 1999-03-18  
; PRIOR APPLICATION NUMBER: 60/078,566  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/078,576  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/078,573  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/078,574  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/078,579  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/080,314  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/080,312  
; PRIOR FILING DATE: 1998-04-01  
; PRIOR APPLICATION NUMBER: 60/078,578  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/078,581  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/078,577  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/078,563  
; PRIOR FILING DATE: 1998-03-19  
; PRIOR APPLICATION NUMBER: 60/080,313  
; PRIOR FILING DATE: 1998-04-01  
; NUMBER OF SEQ ID NOS: 470  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 26  
; LENGTH: 1751  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (1520)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (1557)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (1689)

; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (1729)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (1735)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: SITE  
; LOCATION: (1741)  
; OTHER INFORMATION: n equals a,t,g, or c  
; US-09-397-945-26

Alignment Scores:  
Pred. No.: 8.57e-221 Length: 1751  
Score: 1747.00 Matches: 314  
Percent Similarity: 99.37% Conservative: 0  
Best Local Similarity: 99.37% Mismatches: 2  
Query Match: 99.37% Indels: 0  
Gaps: 0  
DB: 13

US-10-010-050a-2\_copy\_31\_346 (1-316) x US-09-397-945-26 (1-1751)

Qy 1 ArgValSerGlyIleProSerArgArgHisTyrProValProTyrLysArgPheAspPhe 20  
Db 139 CGGGTCTGGGGCATCCCTCCCGGCGCACTGGCGGCTTACAGCGCTTACACTTC 198  
Qy 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40  
Db 199 GCTCCAAACCGAACCTTATGTCAAGTAAATGTAATCTTGTCCCACTGCTCACCT 258  
Qy 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTyr 60  
Db 259 ATCCCAAGTATAGAGGGTGATGATGACATGTAAGATTTTGCATACACCCCGATATG 318  
Qy 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 80  
Db 319 GAATTTAATATGAGAACCTCTCGGACACCTTGAAATATGATGATGATGATGATGATTC 378  
Qy 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTyrGluLeuPheGluLeuGly 100  
Db 379 AGAAGTACCTTAACCTGCAAGAACTACACAAATGATGATGATGATGATGATGATGATG 438  
Qy 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheThrCysAsnGlnGly 120  
Db 439 AACTGTACATTTCCCATCTCCGACCTGAAATGAGATCCCTTCTGGGTATATCAAGGC 498  
Qy 121 AlaAlaCysPhePheGluGlyIleAspAspValHisTyrLysGluAsnGlyThrLeuVal 140  
Db 499 GCTGCTGCTTTTATAGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 558  
Qy 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGluMetAlaLysTyrValLysGlnAsp 160  
Db 559 CAAGTACCACTATACAGAAACATGTTCAACCAATGGCAAAATGGGTGAACAGGAC 618  
Qy 161 AsnGluThrGlyLysTyrTyrGluTyrPheValLysAlaSerProGluLysGlyAla 180  
Db 619 AATGAAACAGGAATTTATATGACATGGAATGTAATGAAACCGACCCGAAAGGGGGCA 678  
Qy 181 GluThrTyrPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200  
Db 679 GAGACATGTTTATCTTACCTTACGACTGTCCAAATTTGTGTTAAGACCTTTAACAAGTTG 738  
Qy 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220  
Db 739 GCTGAATTTGAGAGAGATTCAGAAACATGAAACCACTATACAGAAATATTTCTTAC 798  
Qy 221 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240  
Db 799 AGTGGAGAACCTACTATCTGGAATGAAACATCGTTTTGGGCCAACAGAAACAAG 858  
Qy 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260  
Db 859 ACTCTTGTTTGGCCATTAAGATTTTATTTACCCCTTAACACCATTTGGCAACTAA 918

QY 261 GIUPELEULESERLEULEGNIILEPHEAPALAVALLILEVALHISYGLINPHELYR 280  
DB 919 GAATTCCTGTTGAGCTCTTGCAAAATTTTGTATGACAGATTGTGCACAAACAGTTTCAT 978  
QY 281 LEUPHELYTAAASPHEGLUITYRTTRPHELEUPROMETLYSPHEPROPHETILELYSILETHR 300  
DB 979 TTTGTTTAAATTAATTTGAAATATGGTTTAACTTATGAATTCCTTTTATTAATAATACA 1038  
QY 301 TYRGLUGNIILEPROLEUPROILEARGASNYSTRHLEUSERGLYLEU 316  
DB 1039 TATGAAGAAATCCCTTTACCTATCAGAAACAAACACCTCTGTTTA 1086  
RESULT 5  
US-09-122-383-13  
/ Sequence 13, Application US/09122383A  
/ Patent No. US20020042093A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Sheppard, Paul O.  
/ APPLICANT: Gilbertson, Debra G.  
/ TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
/ TITLE OF INVENTION: CHROMOSOME 13  
/ FILE REFERENCE: 97-38  
/ CURRENT APPLICATION NUMBER: US/09/122,383A  
/ CURRENT FILING DATE: 1998-07-24  
/ EARLIER APPLICATION NUMBER: 60/053,613  
/ EARLIER FILING DATE: 1997-07-24  
/ NUMBER OF SEQ ID NOS: 19  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 13  
/ LENGTH: 1038  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Degenerate nucleotide sequence encoding zs1946  
/ OTHER INFORMATION: polypeptide of SEQ ID NO:2  
/ NAME/KEY: variation  
/ LOCATION: (1)...(1038)  
/ OTHER INFORMATION: n is any nucleotide  
US-09-122-383-13  
Alignment Scores:  
Pred. No.: 2,276-190 Length: 1038  
Score: 1515.00 Matches: 267  
Percent Similarity: 85.03% Conservative: 47  
Best Local Similarity: 85.03% Mismatches: 47  
Query Match: 86.18% Indels: 0  
DB: Gaps: 0  
US-10-010-050A-2\_COPY\_31\_346 (1-316) x US-09-122-383-13 (1-1038)  
QY 2 ValSerGlyIleProSerArgHisTrpProValProTyrIleArgPheAspPheArg 21  
DB 94 GTTWSNGNATHTCCNWSNMGMNCAYTGCCGCTGCTNCAATTAABAGNTTYGATTTWGN .53  
QY 22 ProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerProIle 41  
DB 154 CCAACACCAACGACCTATGTGCARGCAATATAYACNTTYGYCCNACGNGWNSCCNATH 213  
QY 42 ProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValITTPGlu 61  
DB 214 CCGGTATATGARGGNGAYGAYATHTGARGTNTTTCGTTTCCARGCCNCGTNTTGGAR 273  
QY 62 PheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArg 81  
DB 274 TTTAAARAYGNGNAYTNTYNGNCAYTTAAARATHTATGCAYGAYGNCATHTGNTTWMN 333  
QY 82 SerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsn 101  
DB 334 WSNACNTTACNGGNAARAAAYTAAACNATGARGGAYGARGYNTTYCARYTNGNAAY 393  
QY 102 CysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAla 121

DB 394 TGYACNTTYCCNCAAYTTMNGNCCNGARATGAGYGCNCCNTTYTGTGTAAACARGNGCN 453  
QY 122 AlaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuValGln 141  
DB 454 GCNTGYTTTYTGYARGNATGAAGAYGTCNAYTGGAARGAARAAAGNACNTTNTNCR 513  
QY 142 ValAlaThrIleSerGlyAspMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsn 161  
DB 514 GTTNGACNATHTWSNGNAAATATGTTTAAATTCARATGCGNAARTGGGTAAARCARAGAYAN 573  
QY 162 GluThrGlyIleTyrTyrGluThrTrpAsnValLysAspProGluLysGlyAlaGlu 181  
DB 574 GARGACNAGNATHTATATGACACNTGCAAYGTNARGCNSCCNGAARARGNGCNGAR 633  
QY 182 ThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeuAla 201  
DB 634 ACNTGTYGAYWSNTATGAATGYGWSNARTTYGTNTMGNACNTTAAATATATATNGCN 693  
QY 202 GluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSer 221  
DB 694 GARTTYGNGCNGAATTTAARAAATHTGARGACNAAATATACMGNAATHTTNTTATATWSN 753  
QY 222 GlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThr 241  
DB 754 GGNARGCCNACNTATYTTNGNAAAYGACACNWSNGTNTTYGNCNACNAGNAAATARAACN 813  
QY 242 LeuGlyLeuAlaIleLysArgPheTyrTrpProPheLysProHisLeuProThrLysGlu 261  
DB 814 YTMGNNTYNGNATTAARAGNTTYTATATCCNTTYAACCAATYTNCCNACNARARG 873  
QY 262 PheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeu 281  
DB 874 TTYTNTYTWKSNYNTYTNCARATHTTGYAGGNGNATHTGNCAYARCAFTTYATYTN 933  
QY 282 PheTyrAsnPheGluTyrTrpPheLeuProMetLysPheProPheIleLysIleThrTyr 301  
DB 934 TTYTAAATTTGARGATATGTTTATCCNATGAARTTYCCNTTYATTAARATHTACNTAT 993  
QY 302 GluGluIleProLeuProIleArgAsnLysThrLeuSerGly 315  
DB 994 GARGARATHTCCNNTYTNCCNATHTMGNAAATARAACNTYTNWSNGN 1035  
RESULT 6  
US-10-010-050A-13  
/ Sequence 13, Application US/10010050A  
/ Publication No. US20020173624A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Sheppard, Paul O.  
/ APPLICANT: Gilbertson, Debra G.  
/ TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
/ TITLE OF INVENTION: CHROMOSOME 13  
/ FILE REFERENCE: 97-38C1  
/ CURRENT APPLICATION NUMBER: US/10/010,050A  
/ CURRENT FILING DATE: 2002-03-26  
/ PRIOR APPLICATION NUMBER: US 09/122,383  
/ PRIOR FILING DATE: 1998-07-24  
/ PRIOR APPLICATION NUMBER: US 60/053,613  
/ NUMBER OF SEQ ID NOS: 19  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 13  
/ LENGTH: 1038  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: Degenerate nucleotide sequence encoding zs1946  
/ OTHER INFORMATION: polypeptide of SEQ ID NO:2  
/ NAME/KEY: misc\_feature  
/ LOCATION: (1)...(1038)  
/ OTHER INFORMATION: n = A,T,C or G  
US-10-010-050A-13





## Alignment Scores:

Pred. No.:	2,076-107	Length:	494
Score:	886.00	Matches:	163
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	50.40%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x US-09-864-761-22592 (1-494)

```
QY 154 AAlaYsTpVAllyGlnAspAsnGluThrGlyIleYrYrGluThrTrpAsnVallye 173
Db 3 GCAAGTGGTGAACAGCAATGAAACAGAAATTATATGACATGGAATGTA 62
QY 174 AAlaSerProGluYrGlyAlaGluThrTrpPheAspSerYrAspCySerIyPheVal 193
Db 63 GCCAGCCCAAGAAAGGGGGCAGACATGCTTATCTTACGACTGTTCCAAATTTGTG 122
QY 194 LeuArGrThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 213
Db 123 TTAAAGACCTTTAAACAGTGGCTGAAATTGGAGCAGAGTTCAGAAACATGAAACCAAC 182
QY 214 TyrThrArgIlePheLeuYrSerGlyGluProThrYrLeuGlyAsnGluThrSerVal 233
Db 183 TATACAGAAATATTTCTTTACAGTGGAGAACCTTATCTGGAAATGAAACATCTGTT 242
QY 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheYrYrProPhe 253
Db 243 TTTGGGCCAAGAGAAACAGACCTTGGTTTACCAATTAATAATTTTATTTACCCCTTC 302
QY 254 LysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePheAspAlaVal 273
Db 303 AAACCACTTTGCCACTAAAGAAATTTCTGTGAGTCTCTTCAAAATTTTGATGCAAGTG 362
QY 274 IleValHisLysGlnPheYrLeuPheYrAsnPheGluYrTrpPheLeuProMetLys 293
Db 363 ATGTGCACAAACAGTCTTATTTGTTTATATATTTGAATATGTTTATACATGATAA 422
QY 294 PheProHeIleLysIleThrYrGluGluIleProLeuProIleAsnLysThrLeu 313
Db 423 TTCCTTTTATTAATAATACATATGAAAGAAATCCCTTACCTATCAGAAACAAACACTC 482
QY 314 SerGlyLeu 316
Db 483 TCTGCTTA 491
```

## RESULT 8

US-10-040-739-825  
Sequence 825, Application US/10040739  
Publication No. US20020173635A1

## GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John

Lavallee, Edward

Racie, Lisa

Merberg, David

Treacy, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS

NUMBER OF SEQUENCES: 1519

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A

ZIP: 02140

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/040,739

FILING DATE: 07-Jan-2002

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/036,520

FILING DATE: 03-JUN-1998

ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.

REGISTRATION NUMBER: 32,724

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 825:

SEQUENCE CHARACTERISTICS:

LENGTH: 506 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 825:

US-10-040-739-825

## Alignment Scores:

Pred. No.:	2,736-86	Length:	506
Score:	727.00	Matches:	133
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	41.35%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x US-10-040-739-825 (1-506)

```
QY 121 AAlaCyPhePheGluGlyIleAspAsnValHisTrpLysGluAsnGlyThrLeuVal 140
Db 99 GCTGCCCTGCTTTTGGAGGAATGATGATGATTCACAGGAAGAAATGGACATTAAGTT 158
QY 141 GluValAlaThrIleSerGlyAsnMetPheAsnGluMetAlaIleSTrPValLysGluAsp 160
Db 159 CAGTGACCACTATATCAGGAAACATGTTCAACCAATGGCAAGTGGAGAAACAGAC 218
QY 161 AsnGluThrGlyIleYrYrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
Db 219 AATGAAACAGGAATTTATATGAGACATGGAATGTAAACCAACCCAGAAAGGGGCA 278
QY 219 AATGAAACAGGAATTTATATGAGACATGGAATGTAAACCAACCCAGAAAGGGGCA 278
Db 181 GluThrTrpPheAspSerYrAspCySerIyPheValLeuArgThrPheAsnLysLeu 200
QY 279 GAGACATGGTTGATTCCTACGACTGTTCCAAATTTGTTAAGACCTTTAACAGATYG 338
Db 201 AAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTrpThrArgIlePheLeuYr 220
QY 339 GCTGAATTTGAGCAGAGTTCAGAAACATGAAACCAACTATACAGAAATATTTCTTAC 398
Db 221 SerGlyLysProThrYrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
QY 399 AGTGAAGAACCTTATCTGGAATGAAATGAAACATCTGTTTGGGCCAACAGAAACAG 458
Db 241 ThrLeuGlyLeuAlaIleLysArgPheYrYrProPhe 253
QY 459 ACTCTGCTTACCATTAATAAGATTTATATACCCCTTC 497
```

## RESULT 9

US-09-864-761-5832

Sequence 5832, Application US/09864761

Patent No. US20020048763A1

## GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wenheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecmca-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

```

CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 5832
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC001226.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
US-09-864-761-5832

Alignment Scores:
Pred. No.: 1,07e-80 Length: 474
Score: 684.50 Matches: 129
Percent Similarity: 86.67% Conservative: 1
Best Local Similarity: 86.00% Mismatches: 1
Query Match: 38.94% Indels: 19
DB: 9 Gaps: 1

US-10-010-050a-2_copy_31_346 (1-316) X US-09-864-761-5832 (1-474)
QY 123 CysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuValGlnVal 142
DB 82 TGTGTTTTTAAA----- 93
QY 143 AlaThrIleSerGlyAsnKetPheAsnGlnMetAlaIleTrpValIleGlnAspAsnGln 162
DB 94 -----CTAGGAACATGTTCAACCAATGCGAAAGTGGGTGAACGACGACATGAA 144

```

```

QY 163 ThrGlyIleTrpTrpGluThrTrpAsnValIleAsnSerProGluIleGlyValGluThr 182
DB 145 ACAGGAATTATTATTATGACATCGAATCTAAAGCCAGCAAGAAAGGCGACAGAC 204
QY 183 TrpPheAspSerTrpAspCysSerIlePheValIleuArgTrpPheAsnIleValGlu 202
DB 205 TGGTTTGATTCCTCAAGCATGTTCCAAATTTGGTTAAAGACCTTTAAACAAGTGGCTGA 264
QY 203 PheGlyAlaGluPheIleAsnIleGluThrTrpTrpThrValGlyIlePheLeuTrpSerGly 222
DB 265 TTGGAGCAGAGGTCACAGAACATAGAAACCACTATACAAAGAAATATTTCTTTAACAGTGA 324
QY 223 GluProThrTrpLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeu 242
DB 325 GAACCTACTTATCTGGGAAATGAAACATCTGTTTGGGCCAACAGAAACAGACTCTT 384
QY 243 GlyLeuAlaIleIleAspGlyPheTrpTrpProPheLysProHisLeuProThrLysGluPhe 262
DB 385 GGTTCAGCCATTAATAAGATTTTATTAACCCCTTCAACCACTTGGCCAACTTAAGAAATTT 444
QY 263 LeuLeuSerLeuLeuGlnIlePheAspAla 272
DB 445 CTGTTAGTCTCTTGCACAAATTTTGATGCA 474

RESULT 10
US-10-264-049-1859
Sequence 1859, Application US/10264049
Publication No. US20040005579A1
GENERAL INFORMATION:
APPLICANT: Biree et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA133P1
CURRENT APPLICATION NUMBER: US/10/264,049
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: PCT/US01/18569
PRIOR FILING DATE: 2001-06-07
PRIOR APPLICATION NUMBER: US 60/209,467
PRIOR FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 4360
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 1859
LENGTH: 497
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (2)..(2)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (175)..(175)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (285)..(285)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (295)..(295)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (313)..(313)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (323)..(324)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (331)..(331)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:

```

```
/ NAME/KEY: misc_feature
/ LOCATION: (395)..(395)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (440)..(440)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (458)..(458)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (464)..(464)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (478)..(478)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (483)..(483)
/ OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-1859
```

```
Alignment Scores:
Pred. No.: 1.85e-52 Length: 497
Score: 472.00 Matches: 88
Percent Similarity: 98.88% Conservative: 0
Best Local Similarity: 98.88% Mismatches: 1
Query Match: 26.85% Indels: 0
DB: 16 Gaps: 0
```

US-10-010-050a-2\_COPY\_31\_346 (1-316) x US-10-264-049-1859 (1-497)

```
QY 228 G1YANGIUTHRSERVALPHG1YPROTHG1YASNLVSTHLEUG1YLEUAl1eLys 247
DB 3 GGAATGAAACATCTGTTTGGGCCAAGGAAACAGCTCTGTTTGAATATAA 52
QY 248 ArpHeTyrrPrProPhelysProHisLeuProThiLysGluPheLeuSerLeu 267
DB 63 AGATTATTATACCCCTTCAACACCACTTGCACCTAAGAAATTCGTGAGCTCTTG 122
QY 268 G1n1lePheAspAlaVal1leVal1leLysG1nPhetYrLeuPheTyrrAsnPhetYr 287
DB 123 CAATTTTGTATGAGCATTTGTGCACAAACAGTTCTTATTATTATTTGAATAT 182
QY 288 TrpPheLeuProCetyLysPheProPhe1leLys1leThrTyrgluGlu1leProLeuPro 307
DB 183 TGGTTTTCCTATGAAATTCCTTTTATTAAATATACATATGAGAAATCCCTTACT 242
QY 308 1leArpAsnLysThrLeuSerGlyLeu 316
DB 243 ATCAGAAACAAACACCTCTGTGTTTA 269
```

```
RESULT 11
US-09-867-701-2579
/ Sequence 2579 Application US/09867701
/ Patent No. US20020132237A1
/ GENERAL INFORMATION:
/ APPLICANT: Agiate, Paul A.
/ APPLICANT: Jones, Robert
/ APPLICANT: Harlocker, Susan L.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
/ FILE REFERENCE: 210121.497
/ CURRENT APPLICATION NUMBER: US/09/867, 701
/ NUMBER OF SEQ ID NOS: 10912
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2579
/ LENGTH: 351
/ TYPE: DNA
```

ORGANISM: Homo sapien  
US-09-867-701-2579

```
Alignment Scores:
Pred. No.: 9e-38 Length: 351
Score: 359.50 Matches: 78
Percent Similarity: 68.25% Conservative: 8
Best Local Similarity: 61.90% Mismatches: 25
Query Match: 20.45% Indels: 16
DB: 9 Gaps: 3
```

US-10-010-050a-2\_COPY\_31\_346 (1-316) x US-09-867-701-2579 (1-351)

```
QY 82 SerThrLeuThrGlyLysAsnLysThrMetGluThrTyrgluLeuPheGluLeuGlyAsn 101
DB 3 AGTACATTAACTGACAGAACTACACATGAAATGATATGACTTTCCACTTGCGAAC 62
QY 102 CysThrPheProHisLeuArgProGluMetAspAlaProPheTyrrAsnGlnGlyAla 121
DB 63 TGATCATTTCCCGATCTCCGACCTGAAATGATGATCCCTTCTGTGTAATCAAGCGCT 122
QY 122 AlaCysPhePheGluGly1leAspAspValHisThrLysGluAsnGlyThrLeuValGln 141
DB 123 GC-TGCTTTTGGAGAAATTTGATGATGTCACCTGAAAGAAATGGAACATTGATCAA 181
QY 142 Val1leThr1leSerGlyAsnMet-----PheAsnGlnMetAlaLysThrValLysGln 159
DB 182 GTAGCAACTATATACAGTAAAGTTGTGTGAAATATATGCAATTTTATGATGATGATCAACAAA 241
QY 160 AspAsnGluThrGlyLeuTyrrTyrgluThrTrpAsnValLysAlaSerProGluLysGly 179
DB 242 CCAATGAAAGAAATTTGTTAT-----ACTTCC 268
QY 180 AlaGluThrTrpPheAspSerTyrrAspCysSerLysPheValLeuArgThrPheAsn--- 198
DB 269 ATGGAACATTTAGATATGTTTACTTATGAGGTCAATTTGTAATGTACTTTTGAACC 328
QY 199 ---LysLeuAlaGluPhe 203
DB 329 ATTAACCTTTGGGAATTT 346
```

```
RESULT 12
US-09-954-456-1363/C
/ Sequence 1363 Application US/09954456
/ Patent No. US20020115057A1
/ GENERAL INFORMATION:
/ APPLICANT: Young, Paul
/ TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cance
/ TITLE OF INVENTION: Sets
/ FILE REFERENCE: 689290-76
/ CURRENT APPLICATION NUMBER: US/09/954,456
/ PRIOR FILING DATE: 2001-09-18
/ PRIOR APPLICATION NUMBER: US/60/233,617
/ PRIOR FILING DATE: 2000-09-18
/ PRIOR APPLICATION NUMBER: US/60/234,052
/ PRIOR FILING DATE: 2000-09-20
/ PRIOR APPLICATION NUMBER: US/60/234,923
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US/60/235,134
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: US/60/235,637
/ PRIOR FILING DATE: 2000-09-26
/ PRIOR APPLICATION NUMBER: US/60/235,638
/ PRIOR FILING DATE: 2000-09-26
/ PRIOR APPLICATION NUMBER: US/60/235,711
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: US/60/235,720
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: US/60/235,840
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: US/60/235,863
/ PRIOR FILING DATE: 2000-09-27
/ NUMBER OF SEQ ID NOS: 2276
```



```

Db      1164 TCACCGGATCTCTCCGGGAGAAAGCCAAACAGCCGCTTCCCTCTCG-----T 1111
QY      159 InApAsnGluThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysG 1179
Db      1110 CACGACGAGCCAGCGGTTCCAGCAGAGGCTGTGGCAACATGTCATCCGCCAT---G 1054
QY      179 lAlaGluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnL 1099
Db      1053 GTGMAATGGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1001
QY      199 yAlaGluThrGlyAlaGluThrPheLysAsnIleGluThrAsnTyrThrArgIlePheL 219
Db      1001 ----- 1001
QY      219 eutyrserygluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyA 239
Db      1000 -----GCTGGAC 994
QY      239 snLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeuPro 259
Db      993 CCTGCTCTCTGGCGTGAACCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 934
QY      259 hrlYsgluPhe-----LeuIeuSerLeuLeuGlnIlePheAspAlaValIleV 275
Db      933 GCCAGTAATTTCTCTCTTTACTTCTGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 674
QY      275 aHISLysGlnPheTyrLeuPheTyrAsnPhe 285
Db      873 TCTGCAAGCATTTATTCTTTTCCAGTCTTTT 842

```

## RESULT 15

```

US-10-369-493-36401
; Sequence 36401, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 36401
; LENGTH: 986
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-10-369-493-36401

```

## Alignment Scores:

```

Pred. No.: 0.0287 Length: 986
Score: 97.50 Matches: 41
Percent Similarity: 38.22% Conservative: 19
Best Local Similarity: 26.11% Mismatches: 50
Query Match: 5.55% Indels: 47
DB: 16 Gaps: 7

```

JS-10-010-050a-2\_COPY\_31\_346 (1-316) x US-10-369-493-36401 (1-986)

```

JY      138 ThrLeuValGlnValAlaThrIleSeryLysMetPheAsnGlnMetAla----- 154
Jb      168 AGGTTGGTGTGATCGCGGCTTCCAAAGCAACACACGCGGCAAGTGGCGGCGTGGCG 227
JY      155 -----LysTrpValLysGlnAspAsnGlu 162
Jb      228 CGAAGCTTGGCTGCAAGTTCGGCTTGTACAGAAATAATGAGTTGATGGAGCAGCAA 287

```

```

QY      163 ThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGluThr 182
Db      288 AATGGCGTTTACGACAGAGCGGGAATATCCAGCTTTCGCGGCTGATGGCGCAGACAG 347
QY      183 TrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeuAlaGlu 202
Db      348 AGACTCGAT-----ATGTCGCGC 365
QY      203 PheGlyAlaGlu-----PheLysAsnIleGluThrAsnTyrThrArgIlePhe 218
Db      366 TTCGGATTTGAACATAAATTGACCCCTCAAGACTTGAG-----GAGAAATCAGA 416
QY      219 LeuTyrSeryGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
Db      417 GCCAGGCGCGAAAGCCATATTATATC-----CCTGACAGAGCATCATGAT 461
QY      239 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeuPro 258
Db      462 CACCATTTGGCGGCTTAGGCTTCCGACG-----TGGCGGTTTGAAGTTGAGATGACAG 515
QY      259 ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleVal 275
Db      516 GAGAAAGAG-----CTCGGTGATTTCTTGACACTGTGATGTG 554

```

Search completed: April 25, 2004, 06:14:09  
Job time : 636.806 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:02:52 ; Search time 104.894 Seconds  
(without alignment)  
1671.819 Million cell updates/sec

Title: US-10-010-050a-2\_COPY\_31\_346  
Sequence: 1 RVSGIPSRHMPVYKRRDF.....IKTYEELPIPKTKTISGL 316

## Scoring table:

Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODEL=frame+p2n.model -DRV=xih  
-O=/cgn2\_1/USFTO.spool/US10010050/runat.22042004\_113205\_27569/app\_query.fast1\_1.1372  
-DB=Issued\_Patents\_NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOFCI=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10010050 @CGN 1.1 140 @runat.22042004\_113205\_27569 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOBJERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEFY TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued Patents NA:\*  
1: /cgn2\_6/pcodata/2/ina/5A COMB.seq:\*  
2: /cgn2\_6/pcodata/2/ina/5B COMB.seq:\*  
3: /cgn2\_6/pcodata/2/ina/6A COMB.seq:\*  
4: /cgn2\_6/pcodata/2/ina/6B COMB.seq:\*  
5: /cgn2\_6/pcodata/2/ina/PCUTS COMB.seq:\*  
6: /cgn2\_6/pcodata/2/ina/backfile1seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94.5	5.4	14070	4	US-09-108-006C-2
2	91.5	5.2	7718	4	US-09-976-594-244
3	90.5	5.1	1349	1	US-08-290-448A-73
4	90.5	5.1	1349	1	US-08-290-448A-73
5	90.5	5.1	1349	1	US-08-290-448A-73
6	90.5	5.1	1349	1	US-08-175-069A-73
7	90.5	5.1	1349	4	US-08-461-939B-73
8	89.5	5.1	1328	4	US-08-464-000-73
9	89.5	5.1	1328	1	US-08-290-448A-58
10	89.5	5.1	1328	1	US-08-290-448A-58
11	89.5	5.1	1328	1	US-08-175-069A-58
12	89.5	5.1	1328	4	US-08-461-939B-58
					Sequence 58, Appl

13	88.5	5.0	4440	3	US-07-792-600-1	Sequence 1, Appl
14	88.5	5.0	4440	3	US-09-157-021-1	Sequence 1, Appl
15	88.5	5.0	4440	3	US-09-156-842-1	Sequence 1, Appl
16	88.5	5.0	4440	4	US-09-591-514-1	Sequence 1, Appl
17	88.5	5.0	5433	3	US-09-157-021-35	Sequence 35, Appl
18	88.5	5.0	5433	3	US-09-156-842-35	Sequence 35, Appl
19	88.5	5.0	5433	4	US-08-023-655-1363	Sequence 1363, Ap
20	88.5	5.0	5433	4	US-09-591-514-15	Sequence 15, Appl
21	88.5	5.0	1827	4	US-09-620-312D-253	Sequence 253, Appl
22	88.5	5.0	1827	4	US-09-489-039A-6332	Sequence 6332, Ap
23	87.5	5.0	537	4	US-09-134-000C-219	Sequence 219, App
24	86.5	4.9	13930	4	US-09-976-594-1011	Sequence 1011, Ap
25	86.5	4.9	13992	4	US-08-220-132-70	Sequence 70, Appl
26	85.5	4.9	3155	2	US-08-591-629-7	Sequence 7, Appl
27	85	4.8	1830121	4	US-09-557-884-1	Sequence 1, Appl
28	85	4.8	1830121	4	US-09-643-990A-1	Sequence 1, Appl
29	84	4.8	3629	1	US-08-354-618-1	Sequence 1, Appl
30	83	4.7	1116	4	US-09-252-991A-3059	Sequence 3059, Ap
31	83	4.7	1668	4	US-09-134-078-10	Sequence 10, Appl
32	83	4.7	9139	4	US-09-332-478-22	Sequence 22, Appl
33	83	4.7	1664976	4	US-08-916-421B-1	Sequence 1, Appl
34	81.5	4.6	1332	4	US-09-540-236-1387	Sequence 1387, Ap
35	81	4.6	2226	4	US-09-489-039A-683	Sequence 683, App
36	80.5	4.6	2157	4	US-09-328-352-514	Sequence 514, App
37	80.5	4.6	4643	4	US-09-453-702B-72	Sequence 22, Appl
38	80	4.6	1356	4	US-09-489-039A-4678	Sequence 4678, Ap
39	80	4.6	1821	4	US-09-537-682-2	Sequence 2, Appl
40	80	4.6	1959	4	US-09-489-039A-5290	Sequence 5290, Ap
41	79.5	4.5	507	3	US-08-384-106A-3	Sequence 3, Appl
42	79.5	4.5	507	5	PCT-US96-01643-3	Sequence 3, Appl
43	79.5	4.5	594	4	US-09-489-039A-1146	Sequence 1146, Ap
44	79.5	4.5	694	2	US-08-627-610-9	Sequence 9, Appl
45	79.5	4.5	694	3	US-08-384-106A-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-09-108-006C-2  
; Sequence 2, Application US/09108006C  
; Patent No. 6524613  
; GENERAL INFORMATION:  
; APPLICANT: Steer, Clifford J.  
; Kren, Betsy T.  
; Bandyopadhyay, Paramita  
; Roy-Chowdhury, Jayanta  
; TITLE OF INVENTION: Hepatocellular Chimeraplasty  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kimeragen, Inc.  
; STREET: 300 Pheasant Run  
; CITY: Newtown  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 18940  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: IBM Compatible  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/108, 006C  
; FILING DATE: 30-Jun-1992  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/054,288  
; FILING DATE: 30-APR-1997  
; APPLICATION NUMBER: 60/054,837  
; FILING DATE: 05-AUG-1997  
; APPLICATION NUMBER: 60/064,996  
; FILING DATE: 10-NOV-1997  
; APPLICATION NUMBER: 60/074,497  
; FILING DATE: 12-FEB-1998

```
/ APPLICATION NUMBER: PCT US 98/08834
/ FILING DATE: 30-Apr-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Filebel, Thomas
/ REGISTRATION NUMBER: 29258
/ REFERENCE/DOCKET NUMBER: 7991-015-999
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 215-504-4444
/ TELEFAX: 215-504-4545
/
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 14070 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-108-006C-2

Alignment Scores:
Pred. No.: 1.44 Length: 14070
Score: 94.50 Matches: 65
Percent Similarity: 34.57% Conservative: 47
Best Local Similarity: 20.06% Mismatches: 122
Query Match: 5.38% Indels: 90
DB: 4 Gaps: 13

US-10-010-050A-2_COPY_31_346 (1-316) x US-09-108-006C-2 (1-14070)

QY 2 ValSerGlyIleProSerArgArgHisIleProValProTyrylValArgPheAsp-PheAr 21
DB 12075 CTCAGCAGCCTCCCGACCGGAGGAC-----CGGGCGATGATAT 12116
QY 21 GProlysProAspProTyryCysGlnAlaIleTyThrPheCysProThrglySerProI 41
DB 12117 GATGATAGATGACGACCTTTCTTAATGAGACTTCTACTACAGCCCTCTCTTCCA-- 12174
QY 41 eProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValITPQI 61
DB 12175 -----GATTAATAAACTCCACCATATTCATAAAGTGAAGGATGCGCGGA 12218
QY 61 uPheTyryGlyAspLeuLeuGlyHisLeuTyIleMetHisAspAlaIleGlyPheAr 81
DB 12219 A-----TCTGATGAGAACTCAGATCAAGTT----- 12246
QY 81 gSerThrLeuThrGlyAsnTyThrMetGluITrTyrgluLeuPheGlnLeuGlyAs 101
DB 12247 -----AATTGGAAAGAAAGGACGCTTCTGCTT 12275
QY 101 nCyThrPheProHisLeuArgProGluMetAspAlaProPheITrPcyAsnGlnGlyAl 121
DB 12276 GCTAAC-----TCTCTGAAGAACAAGTCCC-----AAGGCCAC 12311
QY 121 aAlaCysPhePheGluGlyIleAspAspValHisITrPlysGlnLysnGly---ThrLeuVa 140
DB 12312 AGGGGCTCTTATATATATATGCAACAAGTACCTGGGAACAACAAGGCTCCACCTCGAG 12371
QY 140 lGlnValAlaITrHisSerGlyAsnMetPheAsnGlnMetAlaIleTyThrValIysGlnAs 160
DB 12372 AGAAGTGTCTCAAGAGCTGAAGAATACTGACAGACCAAGCTGAGTGGTTATCA-- 12429
QY 160 pAsnGluITrGlyIleTyryGluITrTPAsnValIleAsnLaserProGluLysGlyAl 180
DB 12430 -----GGGGCCATTAAGGAATAATTGATGATGACAGAGGTTCCAGAAAGGAGC 12479
QY 180 aGluITrTPheAspSerTyraAspCysSerLysPheValLeuArgThrPheAsnLysIle 200
DB 12480 CAGTGGAGCACACGCGGACCTHACAA----- 12504
QY 200 uAlaGluPheGlyAlaGluPheLysAsnIleGluITrAsn---TyThrArgIlePheLe 219
DB 12505 -----GAGTGAAGAGCAAGAGCCCAAGATCTGTACACAGAACTGTTGAC 12548
```

```
QY 219 uTySerGlyGluProThrTyThrLeuGlyAsnGluITrHisSerValPheGlyProThrglyAs 239
DB 12549 TCAGAAAGCCAGCAGCACTTTCCAGGACTCAAGATAAAGTGTGATGCTGTGTACG 12608
QY 239 nLysThrLeuGlyLeuAlaIleLys----- 247
DB 12609 AGTTACTCAAGAAATTCATATGAAAGTCAGACATCTATGATCATCATTTTCT 12668
QY 248 -----ArgPheTyryITrProPheLysProHisLeuProThrlYsgluPheLeu 264
DB 12669 GAACCTCCCGAGATTCCAGTTCGCGGGAACCTGGATATACACTAGGAGGAACCTTGG 12728
QY 264 uSerLeuGlnIlePheAspAlaValIleValHisLysGlnPheTy----- 280
DB 12729 CACTATGTC---ATAGAGGAGTGAAGGACGATCTGTCCAGATTAATGGAAGTCCA 12785
QY 281 -----LeuPheTyraSnPheGluITrTyThrPheLeuProMetLysPhePr 295
DB 12786 TAATGTTTCAAGAAATACTGTTTCTTATTTCCAAAGACTAGTATTAACCTTCTTGA 12845
QY 295 oPheIleLys 298
DB 12846 GTTAAGGAA 12855

RESULT 2
US-09-976-594-244
; Sequence 244, Application US/0976594
; Patent No. 6673549
;
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; INVENTOR: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 244
; LENGTH: 7718
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 322303.15
; NAME/KEY: unsure
; LOCATION: 6359-6362, 6967-7026
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-244

Alignment Scores:
Pred. No.: 1.35 Length: 7718
Score: 91.50 Matches: 71
Percent Similarity: 31.40% Conservative: 37
Best Local Similarity: 20.64% Mismatches: 133
Query Match: 5.20% Indels: 103
DB: 4 Gaps: 15

US-10-010-050A-2_COPY_31_346 (1-316) x US-09-976-594-244 (1-7718)

QY 11 TrpProValProTyryLysArgPheAsp-----PheArgProLysProAspProTyryCys 28
DB 3443 TGGATGGCTCCGGAATCTATCTTTGACAAATCTACAGACCAAGAGCGATGTGTCT 3502
QY 29 GlnAlaLysTyry-----ThrPheCysProThrglySerProIleProValMetGlu 45
DB 3503 TACGAGTATGCTGTGGGAAATCTTCTCTTAAAGTGGGTCTCCATACCAAGAGTAA 3562
QY 46 GlyAspAspAspIle----- 50
DB 3563 ATGATGAGAACTTTTGCAGTGCCTGAGGGAAGCATGAGATGAGAGCTCTGAGTAC 3622
```



```

QY 51 -----GluValPheArgLeuGlnAlaProValTyrGlu----- 61
Db 3623 TCTACTCTGAAATCATCATGATCATGCTGAGCTGCGACAGAGACCCAAAAAGAAAGG 3682
QY 62 -----PheValTyrGlyAspLeuLeuGlnHisLeuValIleMet 74
Db 3683 CCAGAGATTGACAGACTGTGAGAAAACTAGTGATTTGCTT---CAAGCAATGTACAA 3739
QY 75 HisAspAlaIleLeuPheArgSerThrLeuThrGlyValPheAsnTyrThrMetGluTyr 54
Db 3740 CAGAGT-----GGTAAAGACTACATCCCA-----ATC 3766
QY 95 GluLeuPheGlnLeuGlyAsnGlyThrPheProHisLeuArgProGluMetAspAlaPro 114
Db 3767 AATGCATATGACAGAGAAATAGTGGTTTACATCTCACTCTCTCTCTCTCTCTCTCTGAGAC 3826
QY 115 PheTyrPheAsnGlnGlyAlaAlaCysPhePheGlu-----GlyIleAspAspValHis 132
Db 3827 TTCTTCAGAGAAAGTATTTCAAGCTCCAGAGTTTAATTACAGAAAGCTCTGATGATGCACA 3886
QY 133 TrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 152
Db 3887 TATGTAAATGCTTTCAAGTTCAATGAGCTCGAGAAAGATC-----AAAACTTTGAGAA 3940
QY 153 MetAlaLysTyrValLysGlnAspAsnGluTyrGlyIleTyr-----TyrGlu 168
Db 3941 CTTTAA-----CCGAAATGCCACTCTCATTTTGTAGACATCAAGAGCGGAC 3985
QY 169 ThrTyrAsnValLysLeuAspProGluLysGlyAlaGluThrTyrPheAspTyrAsp 188
Db 3986 AGCAGACACTGTGCGCTCTCCCATGCGAGAGCGCTTCACTGAGACGAGCAAGCAAGCC 4045
QY 189 CysSer-----LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGly----- 204
Db 4046 AAGGCGCTCGCTCAAGATTGACTTGAAGTACACAGTAAAGTAAAGAGTGGGCGCTGCT 4105
QY 205 -----AlaGluPheLysAsn 209
Db 4106 GATGTACAGAGGCCAGTTTCTGCCATTTCCAGCTGTGGGACAGTACAGCAAGCAAGCGC 4165
QY 210 IleGluThr----- 212
Db 4166 AGGTTCACCTACAGACACAGCTGAGCTGGAAGAAATCGCGTGTGCTCCCGCCGCCA 4225
QY 213 AsnTyrThrArgIlePheLeuTyrSerGlyGluPro-----ThrTyr 226
Db 4226 GACTACAACTGGGTGCTCTGACTTCAACCCCACTAGAGTTTGACAGAGAGCTT 4285
QY 227 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle 246
Db 4286 ATTTCAGAGACACATGTGTATTTATACCCCAAGAAACTAGCTTTTCCAGTATTATGC 4345
QY 247 LysArgPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuSerLeu 266
Db 4346 ATATATTAAGTTTACACTTATCTTTCATGAGGAGCAGCTGCTTTTGTGATTTTTTTA 4405
QY 267 LeuGlnIlePhe 270
Db 4406 ATAGTGTCTTT 4417

```

RESULT 3  
 IS-08-290-448A-73  
 Sequence 73, Application US/08290448A  
 Patent No. 5675954

GENERAL INFORMATION:  
 APPLICANT: Rogers, Bruce  
 APPLICANT: Klapper, David G.  
 APPLICANT: Rafnar, Thorunn  
 APPLICANT: Kuo, Mei-chang  
 TITLE OF INVENTION: Allergic Proteins From Ragweed and Uses  
 NUMBER OF SEQUENCES: 93  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHIVE & COCKFIELD

```

STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragoras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMT-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-290-448A-73

Alignment Scores:
Pred. No.: 0.108 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 5.15% Indels: 93
Gaps: 16

US-10-010-050A-2_COPY_31_346 (1-316) x US-08-290-448A-73 (1-1349)
QY 35 CysPro-----ThrGlySerProIleProValMetGluLysAsp 47
Db 526 TGTCAGAGAGCATGATTAACTCCAGATGCTCACCAATTTTAAGACAAAGATGAT 585
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeu 67
Db 586 GGTATGCTTAATATGTTGCTGAGTTCACAAATATG----- 624
QY 68 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87
Db 625 ATGACCATGCTGCTCAGTAAAGCTTCCAGATGGCTCGATACACCTCGGACGC 684
QY 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnGlyThrPheProHisLeu 107
Db 685 TCACACGTG-----ACCGTTTCCAACTGCAAAATTC----- 714
QY 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGly 127
Db 715 -----ACCAACACCAATTTGATTTATTTGCTCGGG 744
QY 128 IleAspAspValHisTyrPheLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
Db 745 GCTGATGACACCCCATTAATCAAGATTAAGGCATGCTA-----GCAACGATGATTC 795
QY 148 AsnMetPheAsnIleMetAlaLysTyrValLysGlnAspAsnGluTyrGlyIleTyr--- 166

```

```

Db 796 AACATGTTCCACGATCACTGTTGACCAAGAAAGCCTAGATGATTTGGTTTTCCAA 355
Qy 167 -----TyrGluThrTyrPasnValylsAlaserProgluylsGlyAlaGlu 181
Db 856 GTCGTTAACACAACTACAGACAGATGGGAAACGTAAGCCATCGGTAGTGGGCCCA 315
Qy 182 Thr-----TyrPheAspSerTyrAspCysSerTyrPheValleu 194
Db 916 ACTATACTCAGCCAGGAAACAGATCTTGGCCCCCGATGATATCATCAAG----- 366
Qy 195 ArgThrPheAsnlysluAlaGluPheGly-----AlaGluPheTyrAsnIleGlu 211
Db 967 -----AAAATGCTTTCAGCAGACGCTGTAACGCAACCCAGAGTGAATGCTGGGAA 1020
Qy 212 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThr 231
Db 1021 TCGAGAACAGATAGACGCTTGAATAATGCT-----:053
Qy 232 SerValPheGlyProThrGly---AsnlyslsThrleuGlyLeuAlaIleTyrArgPheTyr 250
Db 1054 GCTATTTTCTCCCATCCGGGTCTGATCCAGTGTAAACCCCTGAGCAAAAAGCAGGATG 1113
Qy 251 TyrProPheTyrProHleuProThrTyrGlyGluPheLeuSerleuGluIlePhe 270
Db 1114 ATTCACGCTGAAACA-----GGAGAAAGCCGCTCTTAAGACTCACTAGTAGTGTCT 1161
Qy 271 AspAlaValIleValIleTyr-----GlnPheTyrLeuPhe 282
Db 1162 GGTGATCTCATGACCATCAAGACACCTGTGTAAGACACCTGCAATTCCTAAGCTTT 1221
Qy 283 TyrAsn-----PheGluTyrTyrPheLeuProMetIlyPheProHleuIle 299
Db 1222 TATAAATATCATTAATACTATTTATTTTATTT-----TTGATATT 1263
Qy 300 ThrTyrGluGluIleProleuProIleArgAsnlyslsThrleu 313
Db 1264 TTTATATGA-----CCATTAACGTTCAAGTACTCTA 1293

RESULT 4
US-08-290-448A-73
; Sequence 73, Application US/08290448A
; Patent No. 5698204
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce
; APPLICANT: Klapper, David G.
; APPLICANT: Ratnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,448A
; FILING DATE: August 15, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/529,951
; FILING DATE: May 29, 1990
; APPLICATION NUMBER: US 07/325,365
; FILING DATE: March 17, 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-018CN

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1323
; US-08-290-448A-73

Alignment Scores:
Pred. No.: 0.108 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 5.15% Indels: 93
DB: Gaps: 16

US-10-010-050a-2_copy_31_346 (1-316) x US-08-290-448A-73 (1-1349)

Qy 35 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47
Db 526 TGTCCAGGAGGATGATTAATGATCCAGATGGTCCACCAATTTTAAACAAAGATGAT 585
Qy 48 AspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheTyrGlyAspLeu 67
Db 586 GGTGATGCTAATAAAGTCTGCTGATCTTCAACAAATATG----- 624
Qy 68 LeuGlyHisLeuIlyslsMetHisAspAlaIleGlyPheArgSerThrleuThrGlyIys 87
Db 625 ATCGACCAATGCTCGCTCAGTAAGCTTCGATGGGCTCGATGATCAACCTCGGACAG 884
Qy 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnleuGlyAsnCysThrPheProHleu 107
Db 685 TCACACGTC-----ACCGTTCCACTGCAAAATTC----- 714
Qy 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnIlyAlaIleCysPhePheGluGly 127
Db 715 -----ACCAACCAATTTGTATTTGCTCGGG 744
Qy 128 IleAspAspValHisTyrPlyslsGluAsnGlyThrleuValGlnValAlaThrIleSerGly 147
Db 745 GCTGATGACACCCATTAATCAAGTAAAGCATGCTA-----GCAACGGTAGCATTC 795
Qy 148 AsnMetPheAsnGlnMetAlaIlyslsTyrValIysGlnAspAsnGluThrGlyIleTyr 166
Db 796 AACATGTTCCACGATCAAGTGTGACCAAGAAATGCTAGATGTGATTTGGGTTTTCCAA 855
Qy 167 -----TyrGluThrTyrPasnValylsAlaserProgluylsGlyAlaGlu 181
Db 856 GTCGTTAACACAACTACAGACAGATGGGAAACGTAAGCCATCGGTAGTGGGCCCA 915
Qy 182 Thr-----TyrPheAspSerTyrAspCysSerTyrPheValleu 194
Db 916 ACTATACTCAGCCAGGAAACAGATCTTGGCCCCCGATGATATCATCAAG----- 966
Qy 195 ArgThrPheAsnlysluAlaGluPheGly-----AlaGluPheTyrAsnIleGlu 211
Db 967 -----AAAATGCTTTCAGCAGACGCTGTAACGCAACCCAGAGTGAATGCTGGGAA 1020
Qy 212 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThr 231
Db 1021 TCGAGAACAGATAGACGCTTGAATAATGCT----- 1053
Qy 232 SerValPheGlyProThrGly---AsnlyslsThrleuGlyLeuAlaIleTyrArgPheTyr 250
Db 1054 GCTATTTTCTCCCATCCGGGTCTGATCCAGTGTAAACCCCTGAGCAAAAAGCAGGATG 1113

```

QY 251 TyrProPheIysProHisIeuProThrIysGluPheIeuLeuSerIeuLeuGlnIlePhe 270  
DB 1114 ATTCAGCTGAACCA-----GGAGAGCGGCTTGAAGACTCACTAGTGGCT 1161  
QY 271 AspAlaValIleValHisIys-----GlnPheTyrIleuPhe 282  
DB 1162 GGTGTACTTCATGCGCATCAAGACACCTTGCTAAGACCTGCGCAATCTTAAGCTTT 1221  
QY 283 TyrAsn-----PheGluTyrTrpPheIeuProMetIysPheProPheIleIysIle 299  
DB 1222 TATTAATATCATTAATAATCTTATTATTATTATT-----TTTGATATT 1263  
QY 300 ThrTyrGluGluIleProIeuProIleArgAsnIysThrIeu 313  
DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

RESULT 5  
US-08-175-069A-73  
; Sequence 73, Application US/08175069A  
; Patent No. 5776761  
; GENERAL INFORMATION:  
; APPLICANT: Rogers, Bruce  
; APPLICANT: Klapper, David G.  
; APPLICANT: Ratnar, Thorunn  
; APPLICANT: Kuo, Wei-chang  
; TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LAHIVE & COCKFIELD, LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02109-1875  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/175,069A  
; FILING DATE: December 29, 1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/529,951  
; FILING DATE: May 29, 1990  
; APPLICATION NUMBER: US 07/325,365  
; FILING DATE: March 17, 1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Amy E. Mandragouras  
; REGISTRATION NUMBER: 36,207  
; REFERENCE/DOCKET NUMBER: IMI-018DV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)227-7400  
; TELEFAX: (617)227-5941  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1349 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1323  
US-08-175-069A-73

Alignment Scores:  
Pred. No.: 0.108  
Score: 90.50  
Percent Similarity: 32.17%  
Best Local Similarity: 21.66%  
Query Match: 5.15%  
DB: 1  
Length: 1349  
Matches: 68  
Conservative: 33  
Mismatch: 120  
Indels: 93  
Gaps: 16

US-10-010-050A-2\_COPY\_31\_346 (1-316) x US-08-175-069A-73 (1-1349)

QY 35 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47  
DB 526 TGTCCAGAGGAGATGATTAACTCAACAGATGGTCCCAATTTTAAAGACAAAGATGAT 585  
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheIysTyrGlyAspIeu 67  
DB 586 GGTGTATGCTAATAATGTTGCTGTGATGTTCAAAATATG----- 624  
QY 68 LeuGlyHisLeuIysIleMetHisAspAlaIleGlyPheArgSerThrIeuThrGlyIys 87  
DB 625 ATCGACCATTTGCTCCGTCAGTAAGGCTTCGATGAGGCTGCTCGATATCAACCTCGGCAGC 684  
QY 88 AsnTyrThrMetGluTyrTyrGluIeuPheGlnLeuGlyAsnCysThrPheProHisIeu 107  
DB 685 TCACACGTG-----ACCGTTCCAACTGCAAAATTC----- 714  
QY 108 ArgProGluMetAspAlaProPheTyrCysAsnGlnIylAlaIleCysPhePheGluGly 127  
DB 715 -----ACCCAACACCAATTTGTATTATTGCTCGCG 744  
QY 128 IleAspAspValHisTrpIysGluAsnGlyThrIeuValGlnValAlaThrIleSerGly 147  
DB 745 GCTGATGACACCCATTATCAAGATTAAGCATCTA-----GCAACGGTACGATTC 795  
QY 148 AsnMetPheAsnGlnMetAlaIysTyrValIysGlnAspAsnGluThrGlyIleTyr 166  
DB 796 AACATGTTCAACGATCAAGTGTGACCAAAAGATGCTTAAGTATGATTTGGTTTTCCTCA 855  
QY 167 -----TyrGluThrTyrAsnValIylAsnIleSerProGluIysGlyAlaGlu 181  
DB 856 GTCGTTAACAAACACTACAGACAGATGGGAAAGTACGATCGGTAGCTCGGCCCA 915  
QY 182 Thr-----TriPheAspSerTyrAspCysSerIylPheValIeu 194  
DB 916 ACTTACTCAGCCAGGAGGACAGATTCCTGCCCCCGATGATCATCATCAG----- 966  
QY 195 ArgThrPheAsnIylsLeuAlaGluPheGly-----AlaGluPheIysAsnIleGlu 211  
DB 967 -----AAAAATGCTTATAGCAGAGACTGTACTGCAACGCAAGTGCATGTCTGGAAAC 1020  
QY 212 ThrAsnTyrThrArgIlePheIeuTyrSerGlyGluProThrTyrIleGlyAsnGluThr 231  
DB 1021 TGGAGAACAGATGAGACTTGCTTGAATAATGCT----- 1053  
QY 232 SerValPheGlyProThrGly-----AsnIylThrIeuGlyLeuAlaIleIylsArgPheTyr 250  
DB 1054 GCTATTTTCTCCCATCCGGGTCTGATCCAGTGTCAACCCCTGAGCAAAAAGCAGGATG 1113  
QY 251 TyrProPheIysProHisIeuProThrIysGluPheIeuLeuSerIeuLeuGlnIlePhe 270  
DB 1114 ATTCAGCTGAACCA-----GGAGAGCGGCTTGAAGACTCACTAGTGGCT 1161  
QY 271 AspAlaValIleValHisIys-----GlnPheTyrIleuPhe 282  
DB 1162 GGTGTACTTCATGCGCATCAAGACACCTTGCTAAGACCTGCGCAATCTTAAGCTTT 1221  
QY 283 TyrAsn-----PheGluTyrTrpPheIeuProMetIysPheProPheIleIysIle 299  
DB 1222 TATTAATATCATTAATAATCTTATTATTATTATT-----TTTGATATT 1263  
QY 300 ThrTyrGluGluIleProIeuProIleArgAsnIysThrIeu 313  
DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

RESULT 6  
US-08-461-939B-73  
; Sequence 73, Application US/08461939B  
; Patent No. 6335019  
; GENERAL INFORMATION:  
; APPLICANT: Rogers, Bruce

```

APPLICANT: Klapner, David G.
APPLICANT: Ratner, Thorum
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Methods For Treating Sensitivity To A
TITLE OF INVENTION: Protein Allergen Using Peptides which include A T Cell Epitope
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent'n Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,939B
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,000
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy B. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CNDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)742-7400
FAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-461-939B-73
Alignment Scores:
Pred. No.: 0.108
Score: 90.50
Percent Similarity: 32.17%
Best Local Similarity: 21.66%
Query Match: 5.15%
Gaps: 4
Length: 1349
Matches: 68
Conservative: 33
Mismatch: 120
Indels: 93
Gaps: 16
US-10-010-050a-2_COPY_31_346 (1-316) x US-08-461-939B-73 (1-1349)
QY 35 CysPro-----ThnglyserProileProvalMetgluglyaap 47
Db 526 TGTCCAGGAGGCGATGATTAGTCCACGATGGTCCACCATTTTAAAGACAAAGTATGAT 585
QY 48 AspAspIleGluValPheArgIleGlnAlaProValTrpGluPheIleTyrGlyAspLeu 67
Db 586 GGTGATGCTATTAATGCTGCTGAGTGTCAAAATATG-----624
QY 68 LeuGluHisLeuIleValMetHisAspAlaIleGlyPheArgSerThrIleuThrglyys 87
Db 625 ATCGACCATGCTCGCTCAGTAAAGGCTCCGATGGGCTGCTGATATACACCTCGGACGC 684
QY 88 AsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisIleu 107

```

```

Db 685 TCACACGTG-----ACGTTTCCAACTGCAATTC-----714
QY 108 ArgProGluMetAspAlaProPheTrpCysAsnGlnIleAlaIleCysPhePheGluGly 127
Db 715 -----ACCAACACCAATTTGATTTATTTCTGGG 744
QY 128 IleAspAspValHisTrpTyrGluAsnGlyThrIleValGlnValAlaThrIleSerGly 147
Db 745 GCTGATGACACCACTTATTCAGATTAAGATGAGTCA-----GCAACGCTAGCATTC 795
QY 148 AsnMetPheAsnGlnMetAlaIleTyrValIleGlnIleAsnGlnIleuThrglyIleTyr--- 166
Db 796 AAGATGTTCAACCATCATCGTTGACCAAGATGCTTGAATGATTTGTTTCTCAA 855
QY 167 -----TyrGluThrTrpAsnValIleAlaSerProGluIleGlyAlaGlu 181
Db 856 GTGTTAACAAACAACATACAGACAGTGGGACAGCAAGCCATCGGTGTACTCGGCCCA 915
QY 182 Thr-----TrpPheAspSerTyrAspCysSerIlePheValLeu 194
Db 916 ACTATACTACGCAAGGAAACAGATTTCTGCCCCGATGATATATCATCAAG-----966
QY 195 ArgThrPheAsnIleValIleGluPheGly-----AlaGluPheIleAsnIleGlu 211
Db 967 -----AAAAATGCTTAGGAGAGACTGATCTGCAACGCAACGATGCTGTGGAAC 1020
QY 212 ThrAsnTyrThrArgIlePheIleuTyrSerGlyGluProThrTyrIleuGlyAsnGluThr 231
Db 1021 TGGAGAACAGATTAAGACATTTGCTTGAATAATGCT-----1053
QY 232 SerValPheGlyProThrGly-----AsnIleThrLeuGlyIleValIleIleValPheTyr 250
Db 1054 GCTATTTTCTCCATCCGGGCTGTGATCCAGTGTCAACCCGAGCAAAAAGAGAGGATG 1113
QY 251 TyrProPheIlePheProHisLeuProThrIleGluPheIleuSerIleuGlnIlePhe 270
Db 1114 ATCCAGCTGGAACCA-----GGAGAAAGCCGTTCTAGATCTACTAGTATGCTT 1161
QY 271 AspAlaValIleValHisIleIleIleIleIleIleIleIleIleIleIleIleIleIle 282
Db 1162 GGTGATCTCTCATGCTGACGAGGACCTGTGTAACACTGCGCAATTCCTAAGCTTT 1221
QY 283 TyrAsn-----PheGluTyrTrpPheLeuProMetIlePheProIleIleIleIle 299
Db 1222 TATATATATCATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1263
QY 300 ThrTyrGluGluIleProLeuProIleArgAsnIleThrLeu 313
Db 1264 TTATATGAA-----CCATTACCTTCAAGTACTCTA 1293
RESULT 7
US-08-464-000-73
Sequence 73, Application US/08464000
Patent No. 6335020
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapner, David G.
APPLICANT: Ratner, Thorum
TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent'n Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,000
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-464-000-73

Alignment Scores:
Pred. No.: 0.108      Length: 1349
Score: 90.50         Matches: 68
Percent Similarity: 32.17%      Conservative: 33
Best Local Similarity: 21.66%   Mismatches: 120
Query Match: 5.15%             Indels: 93
DB: 4                       Gaps: 16

US-10-010-050a-2_copy_31_346 (1-316) x US-08-464-000-73 (1-1349)
QY 35 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47
DB 526 TGTCCAGGAGGCATGATTAAGTCAACGATGCTGCCACCAATTTTAAGCAACAAAGTGAT 585
QY 48 AaPaPileGluValPheArgLeuGlnAlaProValTTrpGluPheLysTrpGlyAspLeu 67
DB 586 GGTGATGCTATAAATGCTGCTGTGTTACAAATATGG----- 624
QY 68 IeugLysIleLeuLysIleMetHisAspAlaIleGlyPheArgSerThrIleuThrGlyLys 87
DB 625 ATCGACCAATGCTGCTCGCTAGTAAGGCTTCGATGGGCTGCTCGATATATCAACCTCGGACGC 684
QY 88 AsnTYrThrMetGluTrpTYrGluLeuPheGlnLeuGlyAsnCysThrPheProHisIleu 107
DB 685 TCACACGCG-----ACCGTTTCCACCTGCACAAATTC----- 714
QY 108 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 127
DB 715 -----ACCAACACCAATTTGATTATTGCTCGGG 744
QY 128 IleAspAspValHisTrpLysGlnGlnGlyThrLeuValGlnValAlaThrIleSerGly 147
DB 745 GCTGATGACACCACTATATCAAGATTAAGGCAATGCTA-----GCAACGCTAGCAATTC 795
QY 148 AsnMetPheAsnGlnMetAlaLysTrpValysGlnAspAsnGluThrGlyIleTYr--- 166
DB 796 AACATGTTTCCACGATCACTGACCAAAAGATGCTAGATGATTTGGGTTTTCCAA 855
QY 167 -----TYrGluThrTrpAsnValLysAlaSerProGluLysGlyValGlu 181
DB 856 GTTCGTTAACAACTACGACGATCGGGAAGTACGCCATGCGGTGAGTCTGGACCCCA 915
QY 182 Thr-----TrpPheAspSerTYrAspCysSerLysPheValLeu 194
DB 111 -----

```

```

DB 916 ACTATACTCAGCAAGGGAACAGATTTCTGCCCCCGCATGATATCATCAG----- 966
QY 195 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 211
DB 967 -----AAAAATGCTTTAGCGAGACTGCTACTGGCAACGAGATGCTGCGGAAC 1020
QY 212 ThrAsnTYrThrArgGlnLeuPheLeuLysSerGlyGluProThrTYrLeuGlyAsnIleuThr 231
DB 1021 TGGAGAACAGATGAGACTTCGTTGAATAATGCT----- 1053
QY 232 SerValPheGlyProThrGly---AsnLysThrLeuGlyLeuAlaIleLysArgPheTYr 250
DB 1054 GGTATTTTTCCTCCATCCGGGCTGATCCAGTGCTAACCCCTGAGCAAAAACAGAGATG 1113
QY 251 TYrProPheLysProHisLeuProThrLysGlnPheLeuSerLeuLeuGlnIlePhe 270
DB 1114 ATTCAGCTGAACCA-----GAGAAAGCGTTCTTAAGACTCACTAGTAGTGCT 1161
QY 271 AspAlaValIleValHisLys-----GlnPheTYrLeuPhe 282
DB 1162 GGTGACTCTCATGCCATCAAGAGCACTTGTAAAGCACTGGGCATTCCTAAAGCTTT 1221
QY 283 TYrAsn-----PheGluTYrTrpPheLeuProMetLysPheProPheIleLysIle 299
DB 1222 TATAATATCATTAATACTTATTTATTTATTT-----TTGATATTT 1263
QY 300 ThrTYrGluGluIleProLeuProIleArgAsnLysThrLeu 313
DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

RESULT 8
US-08-290-448A-58
Sequence 58, Application US/08290448A
Patent No. 5676954
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafner, Thorum
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1328  
US-08-290-448A-58

## Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	0.143	1328	68	28	113	15
Percent Similarity:	89.50					
Best Local Similarity:	29.45%					
Query Match:	20.86%					
DB:	5.09%					

US-10-010-050a-2\_copy\_31\_346 (1-316) x US-08-290-448A-58 (1-1328)

```

QY 35 CysPro-----ThrglySerProIleProValMetGluGlyAsp 47
   |||||
Db 505 TGTCCAGGAGCGCATGATTAGTCCACAGCATGTCACCAATTAAAGACAAAGTGAT 564
   |||||
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTPGluPheLysTYrGlyAspLeu 57
   |||||
Db 565 GGTATGCTAATAAATGTTGCTGCTAGTTCACAAATATG----- 503
   |||||
QY 68 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 37
   |||||
Db 604 ATCCACCATGCTGCTGCTAGTAAGGCTTCGATGGCTGCTGATACACCTCGGAGC 563
   |||||
QY 88 AsnTYrThrMetGluTPYrGluLeuPheGlnLeuGlyLysCysThrPheProHisLeu 107
   |||||
Db 664 TCACACGTG-----ACCGTTTCACATCCCAATTC----- 593
   |||||
QY 108 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaIleCysPhePheGluGly 127
   |||||
Db 694 -----ACCCAAACCAATTTGATTATTCCTCGGG 723
   |||||
QY 128 IleAspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
   |||||
Db 724 GCTGATGACACCCATATCAAGATTAAGGATGCTA-----GCAAGGTAGACATTC 774
   |||||
QY 148 AsnMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGlnThrGlyIleTYr--- 166
   |||||
Db 775 AACATGTTACCCGATCAGTTGACCAAGAAATGCTGATGATTGGGTTTTCCAA 834
   |||||
QY 167 -----TYrGluThrTPAsnValLysAlaSerProGluLysGlyAlaGlu 81
   |||||
Db 835 GTCTTAACAACAACACTACGACAGATGGGAAACGTACGCCATCGGTGTAAGTCCGCCCA 894
   |||||
QY 182 Thr-----TrpPheAspSerTYrAspCysSerLys----- 931
   |||||
Db 895 ACTATACTACGCAAGGAAACAGATTCCTGCCCCGATGATATCATCAAGAAATGTC 954
   |||||
QY 192 -----PheValLeuAlaGlnThrPheAsn 998
   |||||
Db 955 TTACGCGAGACTGGTACTGGCAACGACAGTCATGTCGGGAACCTGGAAACAGATTA 1014
   |||||
QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGlnThrAsnTYrThrArgIlePhe 218
   |||||
Db 1015 GACTGTGCTGAATAATGCTGCT-----ATTTT 2041
   |||||
QY 219 LeuTYrSerGlyGluProThrTYrLeuGlyAsnGlnThrSerValPheGlyProThrGly 238
   |||||
Db 1042 CTCCTCATCCGGGTCTGATCCAGTGTAAACCTGAGACA----- 2080
   |||||
QY 239 AsnLysThrLeuGlyLeuAlaIleLysArgPheTYrTYrProPheLysProHisLeuPro 258
   |||||
Db 1081 -----AAAGACGAGATGATTTCACTGAACCA 1107
   |||||
QY 259 ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLys--- 277
   |||||
Db 1108 ---GAGAAAGCCGTTCTTAAGATCTCACTAGTAGTGCTGTTACTCTCATGCCATCAAGA 1164

```

```

QY 278 -----GlnPheTYrLeuPheTYrAsn-----PheGluTYr 287
   |||||
Db 1165 GCACTTGCTAAGACCTGGCCGATTCCTAAGCTTTATATATATCAATAATACTTATT 1224
   |||||
QY 288 TrpPheLeuProMetLysPheProPheIleLysIleThrTYrGluGluIleProLeuPro 307
   |||||
Db 1225 TATTTTATTT-----TTGATATTTTATATGAA-----CCA 1254
   |||||
QY 308 IleArgAsnLysThrLeu 313
   |||||
Db 1255 TTACGTCAGACTACTCTA 1272
   |||||
RESULT 9
US-08-290-448A-58
: Sequence 58, Application US/08290448A
: Patent No. 5698204
: GENERAL INFORMATION:
: APPLICANT: Rogers, Bruce
: APPLICANT: Klapper, David G.
: APPLICANT: Rafnar, Thorunn
: APPLICANT: Kuo, Mei-chang
: TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
: NUMBER OF SEQUENCES: 93
: CORRESPONDENCE ADDRESS:
: ADDRESS: LAHIVE & COCKFIELD
: STREET: 60 State Street, suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/290,448A
: FILING DATE: August 15, 1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/529,951
: FILING DATE: May 29, 1990
: APPLICATION NUMBER: US 07/325,365
: FILING DATE: March 17, 1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Amy E. Mandragouras
: REGISTRATION NUMBER: 36,207
: REFERENCE/DOCKET NUMBER: IMI-018CN
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 58:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1328 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1328
: US-08-290-448A-58
:
Alignment Scores:
Pred. No.: 0.143 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 5.09% Indels: 117
DB: 1 Gaps: 15
US-10-010-050a-2_copy_31_346 (1-316) x US-08-290-448A-58 (1-1328)
QY 35 CysPro-----ThrglySerProIleProValMetGluGlyAsp 47

```

```

Db      |||||
505  TGTCCAGAGGAGCATGATTAAAGTCCAAAGCATGTCACCAATTTTAAAGCAACAAGTAT 564
Qy      |||||
48  AspAspIleGluValPheArgLeuGlnAlaProValTTPGluPheLeuTyrlAspLeu 67
Db      |||||
555  GGTGATGCTATAATGTTGCTGTAGTCTACAAATATG----- 603
Qy      |||||
68  LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87
Db      |||||
604  ATCGACCAATGCTGCTGCTAGTAAGCTTCGAGGCTGCTCGATATACCTCGGAGC 653
Qy      |||||
88  AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCySerThrPheProHisLeu 137
Db      |||||
664  TCACACGTC-----ACCGTTCCAACTGCAATTC----- 693
Qy      |||||
108  ArgProGluMetAspAlaProPheTyrPysAsnGlnGlyAlaAlaCysPhePheGluGly 127
Db      |||||
694  -----ACCCAAACCAATTTGATTTATTTGCTCGGG 723
Qy      |||||
128  IleAspAspValHisTyrPheGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
Db      |||||
724  GGTATGACACCATATCAAGATTAAGGCATGCTA-----GCAACGGTAGCATTC 774
Qy      |||||
148  AsnMetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluThrGlyIleTyr--- 166
Db      |||||
775  AACATGTTCCACGATCAGCTGACCAAGAATGCTAGATGATGATTTGGGTTTTTCCA 834
Qy      |||||
167  -----TyrGluThrTyrAsnValLysAlaSerProGluLysGlyAlaGlu 181
Db      |||||
835  GTCGTTAAACAACACTACAGACAGATGGGGAACGACGCGATGCTGAGTCTGCGCCCA 894
Qy      |||||
182  Thr-----TyrPheAspSerTyrAspCysSerLys----- 191
Db      |||||
895  ACTATACTCAGCCAGGAGACAGATTTCTCCGCCCATGATATCATCAAGAAAAATGTC 954
Qy      |||||
192  -----PheValLeuArgThrPheAsn 198
Db      |||||
955  TTAGCGAGAGCTGGTACTGCGACAGCAGAGTCAGATGCTGCGAATCTGGAGAAAGATAA 1014
Qy      |||||
199  LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 218
Db      |||||
1015  GACTTGCTTGAATATGCTGCT-----ATTTT 1041
Qy      |||||
219  LeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
Db      |||||
1042  CTCCCATCCGGGTGTATCCAGTGTACCCCTAGCAA----- 1080
Qy      |||||
239  AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuPro 258
Db      |||||
1081  -----AAAGCAGGATGATTCAGCTGAACCA----- 1107
Qy      |||||
259  ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLys--- 277
Db      |||||
1108  ---GAGAAAGCGTTTAAAGTCACTAGTAGTGTGCTGATCTCATGCGCAACAGGA 1164
Qy      |||||
278  -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 287
Db      |||||
1165  GCACCTTGCTTAAGACCTCGGCATTTCTAAGCTTTTAAATATCAATAATATTATTT 1224
Qy      |||||
288  TyrPheLeuProMetLysPheProPheIleLysIleThrTyrGluGlnIleProLeuPro 317
Db      |||||
1225  TATATTATTT-----TTGATATTTTATATGAA-----CCA 1254
Qy      |||||
308  IleArgAsnLysThrLeu 313
Db      |||||
1255  TTACGTTCAAGTACTCTA 1272

```

RESULT 10  
 US-08-175-069A-58  
 ; Sequence 58, Application US/08175069A  
 ; Patent No. 576761  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rogers, Bruce

```

; APPLICANT: Klapper, David G.
; APPLICANT: Rafnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Allergenic Proteins from Ragweed and Uses
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESS: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/175,069A
; FILING DATE: December 29, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/529,951
; FILING DATE: May 29, 1990
; APPLICATION NUMBER: US 07/325,365
; FILING DATE: March 17, 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy B. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-018DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1328
; US-08-175-069A-58

Alignment Scores:
Pred. No.: 0.143 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 5.09% Indels: 117
DB: 1 Gaps: 15
US-10-010-050A-2_COPY_31_346 (1-316) x US-08-175-069A-58 (1-1328)

Qy      |||||
35  CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47
Db      |||||
505  TGTCCAGAGGAGCATGATTAAAGTCCAAAGCATGTCACCAATTTTAAAGCAACAAGTAT 564
Qy      |||||
48  AspAspIleGluValPheArgLeuGlnAlaProValTTPGluPheLeuTyrlAspLeu 67
Db      |||||
555  GGTGATGCTATAATGTTGCTGTAGTCTACAAATATG----- 603
Qy      |||||
68  LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87
Db      |||||
604  ATCGACCAATGCTGCTGCTAGTAAGCTTCGAGGCTGCTCGATATACCTCGGAGC 653
Qy      |||||
88  AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCySerThrPheProHisLeu 107
Db      |||||
664  TCACACGTC-----ACCGTTCCAACTGCAATTC----- 693
Qy      |||||
108  ArgProGluMetAspAlaProPheTyrPysAsnGlnGlyAlaAlaCysPhePheGluGly 127
Db      |||||
694  -----ACCCAAACCAATTTGATTTATTTGCTCGGG 723

```



QY 128 ILeaspapValH:stRPLySGluAsnGlyThrLeuValGluValAlaThrIleSerGly 147  
 Db 724 GCTATGACCCCATTTTCAGATTAAGCATGCTA-----GCAAGCATGACATTC 774  
 QY 148 AsnMetPheAsnGlnMetAlaIleTyrValLysGlnAspAsnGluThrGlyIleTyr--- 166  
 Db 775 AACATGTTACCCGATCAGCTTGCACCAAGAAATGCTAGATGTAGATTTGGTTTTCGAA 334  
 QY 167 -----TyrGluThrTTPAsnValLysAlaSerProGluLysGlyValGlu 181  
 Db 835 GTGCTTAACACACACTAGACAGATGGGAACGTAACCATCGGTAGTGGTCCCA 394  
 QY 182 Thr-----TTPheAspSerTyrAspCysSerLys----- 191  
 Db 895 ACTATCTACAGCCAGGAAGAAAGATCTTCCGCCCATGATATCATCAAGAAATGTC 354  
 QY 192 -----PheValLeuAlaGlyThrPheAsn 398  
 Db 955 TTAGCAGAGACTGCTACTGCGACAGAGTCAGTGTCTGGAACTGGAAACAGATAAA 1014  
 QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 218  
 Db 1015 GACTTGCTGAAAATGCTGCT-----ATTTT 1041  
 QY 219 LeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238  
 Db 1042 CTCCTCATCGGGTCTGATCCAGTCAACCCCTAGACAA----- 1080  
 QY 239 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeuPro 258  
 Db 1081 -----AAAGCAGGATGATTCACCTAACA 1107  
 QY 259 ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHISLys--- 277  
 Db 1108 ---GAGAAGCCCTTAAAGACTCACTAGAGTGGTGTACTCTCAACCATCAAGAA 1164  
 QY 278 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 287  
 Db 1165 GCACCTTGCTAAGACACCTGCGCAATTCCTAAGCTTTTAAATAAATCAATAATACCTATT 1224  
 QY 288 TTPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 307  
 Db 1225 TMTTTTATT-----TTTGATTAATTTATATGAA-----CCA 1254  
 QY 308 ILeArgAsnLysThrLeu 313  
 Db 1255 TTACGTTCAAGTACTCTA 1272  
 RESULT 11  
 US-08-461-939B-58  
 / Sequence 58, Application US/08461939B  
 / Patent No. 6335019  
 / GENERAL INFORMATION:  
 / APPLICANT: Rogers, Bruce  
 / APPLICANT: Klapper, David G.  
 / APPLICANT: Rafnar, Thorunn  
 / APPLICANT: Kuo, Wei-chang  
 / TITLE OF INVENTION: Methods For Treating Sensitivity To A  
 / NUMBER OF SEQUENCES: 93  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: LAHIVE & COCKFIELD, LLP  
 / STREET: 28 State Street  
 / CITY: Boston  
 / STATE: Massachusetts  
 / COUNTRY: USA  
 / ZIP: 02109-1875  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: PatentIn Release #1.0, Version #1.25

/ CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/461,939B  
 / FILING DATE:  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: US 08/464,000  
 / FILING DATE: 05-JUN-1995  
 / APPLICATION NUMBER: US 08/290,448  
 / FILING DATE: 15-AUG-1994  
 / APPLICATION NUMBER: US 07/529,951  
 / FILING DATE: 29-MAY-1990  
 / APPLICATION NUMBER: US 07/325,365  
 / FILING DATE: 17-MAR-1989  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Amy E. Mandragouras  
 / REGISTRATION NUMBER: 36,207  
 / REFERENCE/DOCKET INFORMATION:  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (617)227-7400  
 / TELEFAX: (617)742-4214  
 / INFORMATION FOR SEQ ID NO: 58:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 1328 base pairs  
 / TYPE: nucleic acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: cDNA  
 / FEATURE:  
 / NAME/KEY: CDS  
 / LOCATION: 1..1328  
 / US-08-461-939B-58  
 Alignment Scores:  
 / Pred. No.: 0.143 Length: 1328  
 / Score: 89.50 Matches: 68  
 / Percent Similarity: 29.45% Conservative: 28  
 / Best Local Similarity: 20.86% Mismatches: 113  
 / Query Match: 5.09% Indels: 117  
 / DB: 4 Gaps: 15  
 US-10-010-050a-2\_copy\_31\_346 (1-316) x US-08-461-939B-58 (1-1328)  
 QY 35 CysPro-----ThrGlySerProIleProValMetGluLysP 47  
 Db 505 TGTCCAGAGAGGATGATTAAGTCCACAGATGCTCCACCAATTTTAAAGACAAAGATGAT 564  
 QY 48 AspAspIleGluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeu 67  
 Db 565 GGTATGCTATTAATGTTGCTGTAGTTCACCAATATG----- 603  
 QY 68 LeuGlyHISLeuLysIleMetHISAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87  
 Db 604 ATTCACCATTTGCTGCTGCTAGTAAGGCTTCCGATGGCTGCTGATATGACCCCTGGCAGC 663  
 QY 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHISLeu 107  
 Db 664 TCACACGCG-----ACCGTTTCCAACTCAATTC----- 693  
 QY 108 ArgProGluMetAspAlaProPheTyrPheAsnGlnGlyAlaLysPhePheGluGly 127  
 Db 694 -----ACCCACACCAATTTGATTTATGCTCGGG 723  
 QY 128 ILeaspapValH:stRPLySGluAsnGlyThrLeuValGluValAlaThrIleSerGly 147  
 Db 724 GCTATGACCCCATTTTCAGATTAAGCATGCTA-----GCAAGCATGACATTC 774  
 QY 148 AsnMetPheAsnGlnMetAlaIleTyrValLysGlnAspAsnGluThrGlyIleTyr--- 166  
 Db 775 AACATGTTACCCGATCAGCTTGCACCAAGAAATGCTAGATGTAGATTTGGTTTTCGAA 834  
 QY 167 -----TyrGluThrTTPAsnValLysAlaSerProGluLysGlyValGlu 181  
 Db 835 GTGCTTAACACACACTAGACAGATGGGAACGTAACCATCGGTAGTGGTCCCA 894

```

QY 182 Thr-----TrrpheaspseryzaspCysSerLys-----191
Db 895 ACTAATACAGCAAGGAAAGAGATCTTCCGCCCGATATATCATCAAGAAATGTC 554
QY 192 -----PheValLeuArgThrPheAsn 198
Db 955 TTACGAGAGACTGTGTAAGTGCACGAGAGTCGATGCGGAACTGAGAACTAA 1014
QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTrpThrArgIlePhe 218
Db 1015 GACTTCTGTAATAATGCTGCT-----ATTGTT 1041
QY 219 LeuTyrSerGlyGluProThrTrpTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
Db 1042 CTCACATCCGGGCTGATCCAGTCAAGCTTAACCCCTGAGCA-----1080
QY 239 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTrpProPheLysProHisLeuPro 258
Db 1081 -----AAGAGAGGATGATCCAGCTGAGCA-----1107
QY 259 ThrLysGluPheLeuLeuSerLeuGlnIlePheAspAlaValIleValHisLys---277
Db 1108 ---GGAGAAAGCGCTTAAAGACTACTAGTACTGCTGCTGATCTCATGCAAGCA 1164
QY 278 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 287
Db 1165 GCACCTTGCTAAGCACTGCGCAATTCCTAAGCTTTATATATATCAATAAATCTTATTT 1224
QY 288 TrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 307
Db 1225 TATTATTATAT-----TTTGATATTATATAGAA-----CCA 1254
QY 308 IleArgAsnLysThrLeu 313
Db 1255 TTACGTTCAAGTACTCTA 1272

RESULT 12
US-08-464-000-58
; Sequence 58, Application US/08464000
; Patent No. 6335020
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce
; APPLICANT: Klapper, David G.
; APPLICANT: Rafnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSER: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,000
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/290,448
; FILING DATE: 15-AUG-1994
; APPLICATION NUMBER: US 07/529,951
; FILING DATE: 29-MAY-1990
; APPLICATION NUMBER: US 07/325,365
; FILING DATE: 17-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-01BCN2

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1328
; US-08-464-000-58

Alignment Scores:
Pred. No.: 0.143 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 5.09% Indels: 117
DB: Gaps: 15

US-10-010-050A-2_COPY_31_346 (1-316) x US-08-464-000-58 (1-1328)
QY 35 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 47
Db 505 TGTCCAGAGAGCATATTAAGTCAACAGATGCTCCACCAATTTTAAGACAAAGATGAT 564
QY 48 AspAspIleGluValPheArgLeuGlnAlaProValITrpGluPheLysTyrGlyAspLeu 67
Db 555 GGTATGCTATAAATGTTGCTGCTGATGTTCACAAAATATG-----603
QY 68 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 87
Db 604 ATGCACCATTCGTGCTCAAGTAAGGCTTCGATGGCTCGATATACCTCGGACG 663
QY 88 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 107
Db 664 TCACACGTG-----ACCGTTTCCACTGCAAAATTC-----693
QY 108 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 127
Db 694 -----ACCAACCACAATTTGATATATGCTCGGG 723
QY 128 IleAspAspValHisTyrLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 147
Db 724 GCTGATGACACCATTTATCAAGATTAAGGATGCTA-----GCAACGTTAGCATTC 774
QY 148 AsnMetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluThrGlyTyr---166
Db 775 AACATGTTCAACCGATCAAGTTCACCAAGATGCTAGATGATTTGGGTTTTCCAA 834
QY 167 -----TyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGlu 181
Db 835 GTGCTTAACAACAACATTCGACAGATGGGGAACGATCCCATCGGTAGCTGGGCCCA 894
QY 182 Thr-----TrrpheaspseryzaspCysSerLys-----191
Db 895 ACTAATACAGCAAGGAAAGAGATCTTCCGCCCGATATATCATCAAGAAATGTC 954
QY 192 -----PheValLeuArgThrPheAsn 198
Db 955 TTACGAGAGACTGTGTAAGTGCACGAGAGTCGATGCGGAACTGAGAACTAA 1014
QY 199 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTrpThrArgIlePhe 218
Db 1015 GACTTCTGTAATAATGCTGCT-----ATTGTT 1041
QY 219 LeuTyrSerGlyGluProThrTrpTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 238
Db 1042 CTCACATCCGGGCTGATCCAGTCAAGCTTAACCCCTGAGCA-----1080

```

QY 239 AenlystThrLeuGlyLeuAlaIleIysArgPheTyrTrpProPheLysProHisLeuPro 258  
Db 1081 -----AAAGCAGGAGTATCCAGCTGAACCA----- 1107  
QY 259 ThrLysGluPheLeuLeuSerLeuGluIlePheAspAlaValIleValHisLys--- 277  
Db 1108 ---GGAGAAAGCGTCTAGAGACTCAGTACGTGCTGCTGACTCTCATGCCATCAAGGA 1164  
QY 278 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 287  
Db 1165 GCACCTTGCTTAAGCAGCTGCGCAATTCCTTAATTAATCAATAATTAATTT 1224  
QY 288 TrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 307  
Db 1225 TATTTTATT-----TTGATATTATTATGAA-----CCA 1254  
QY 308 IleArgAsnLysThrLeu 313  
Db 1255 TTACGTTCAAGTACTCTA 1272  
RESULT 13  
US-07-792-600-1  
Sequence 1, Application US/07792600  
Patent No. 6008045  
GENERAL INFORMATION:  
APPLICANT: COPELAND, WILLIAM C.  
APPLICANT: WANG, TERESA S.-F.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
NUMBER OF INVENTION: TEMPLATE-DEPENDENT ENZYMAIC SYNTHESIS OF NUCLEIC ACID  
CORRESPONDENCE ADDRESS:  
ADDRESS: Peter G. Carroll  
STREET: 220 Montgomery Street, Suite 710  
CITY: San Francisco  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/792,600  
FILING DATE: 19911115  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CARROLL, PETER G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: STDU-00097  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4440 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: Single  
TOPOLOGY: linear  
US-07-792-600-1  
Alignment Scores:  
Pred. No.: 1.37 Length: 4440  
Score: 88.50 Matches: 71  
Percent Similarity: 36.39% Conservative: 56  
Best Local Similarity: 20.34% Mismatches: 115  
Query Match: 5.03% Indels: 107  
Gaps: 20  
US-10-010-050A-2\_COPY\_31\_346 (1-316) x US-07-792-600-1 (1-4440)  
QY 16 LysArgPheAspPheArgProLysProAspProTyrCysGlnAlaIleLysThrPheCys 35  

Db 563 AAGAAAAGATCCATTGGAGGCTTACCGAATCTTTCTCTGACACACCGCAGCAATT 622  
QY 36 ProthrGly-----SerProIlePro 42  
Db 623 CTTTCAGAAAAAATTGCTTCCTCCCTGCTCCAGAAAGAGCTCCATTAACTCGTTCTCT 682  
QY 43 ValMetGluGlyAsp-----AspAspIleGluVal-----PheArgLeu 55  
Db 683 CTTAAAGTGTCTGAATTGCTGCGGATGATGTATACAGGTCCAGACTACAGAAAGAGCAG 742  
QY 56 GlnAlaProValTrpGluPheLysTyrGlyAspLeuLeuGluIleLysLysIleMetHis 75  
Db 743 GAGTCAGGCGCAATGAGATTGAAGATGTGACTTT----- 778  
QY 76 AspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyrTrpGlu 95  
Db 779 -----GATAGCCCATGAGAAATGGAAG 802  
QY 96 LeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGluMetAspAlaProPhe 115  
Db 803 GTG-----GACTGAGGCT--ATGGCTGCCAGGCT 832  
QY 116 TrpCysAsnGlnGlyAlaIleCysPhePheGluIleAspAspValHisTrpLysGlu 135  
Db 833 TGGACAAAGAGAGT-----GAGCCAGACAGAGAGTGAACAAGAGCG 877  
QY 136 Asn-----GlyThrLeuValGlnValAlaIleThrIleSerGlyAsnMetPheAsnGln 152  
Db 878 GATTCGGGAAAGAGGCGGT-----TCTACTTAAGAAATTTTCCCGGAT 925  
QY 153 MetAlaLysTrp---ValLysGlnAspAsnGluThrGlyIleTyrTrpGluThrTrpAsn 171  
Db 926 GTCTCTGTTGGACATGATCAAGAAGTGAATGAGATTTCTCACTGCAAGAAATTCAA 985  
QY 172 ValLysAlaSerPro-----GluTyrGlyAlaGluThr----- 182  
Db 986 GTGATTCACAGTCACTCCCATGTTGTAAGGGGCGAGTAGAGAACAAAGTATCCACTTT 1045  
QY 183 ---TrpPheAspSerTyrAsp-----CysSerLysPheValLeuArgThrPheAsn 198  
Db 1046 TATGTGTTGATGGCTTATAGAGATCAGTACAAACCAACGAGTGTGATTTCTGTGGG 1105  
QY 199 LysLeuAlaGluPheGlyValGlu-----PheLysAsnIle 210  
Db 1106 AAGTTTGATTGAATCAAGCCGAGACCATGTAGCTGTGTCTATGGTGAATAATATC 1165  
QY 211 GluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTrpLeuGlyAsnGlu 230  
Db 1166 GAGCGAAGC-----CTTAC-----TTCCTCCCGTGAA 1195  
QY 231 ThrSerValPheGlyProThrGlyAsnLysThr---LeuGlyLeuAlaIleLysArgPhe 249  
Db 1196 ATGAAATTTGATCTAAATACGGGGAAGAAACAGCAACTCCATTTCAATGAAGATGTT 1255  
QY 250 TyrTyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGluIle 269  
Db 1256 TATGAGCAATTGATGAGAAATAGCAACAAATATAAATTATGAG----- 1303  
QY 270 PheAspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPhe 289  
Db 1304 TTCAAGTCTAAGCCAGTGAAGAAAGAAC-----TATGCTTTGAGATACCTGAT 1351  
QY 290 LeuProMetLysPhePheProPheIleLysIleThrTyrGlu-----GluLeuPro 305  
Db 1352 GTTCCAGAAAATTCGATGACTCTGGAATTAATCTCGCGCTGAATGCCACAGCTTCT 1411  
QY 306 LeuProIleArgAsnLysThrLeuSer 314  
Db 1412 CAAGATTGAAAGAGAAACTTTTCT 1438  
RESULT 14  
US-09-157-021-1  
Sequence 1, Application US/09157021A

```

; Patent No. 6100023
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/157,021A
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-157-021-1

Alignment Scores:
Pred. No.: 1.37      Length: 4440
Score: 88.50      Matches: 71
Percent Similarity: 36.39%      Conservative: 56
Best Local Similarity: 20.34%      Mismatches: 115
Query Match: 5.03%      Indels: 107
DB: 3      Gaps: 20

US-10-010-050a-2_COPY_31_346 (1-316) x US-09-157-021-1 (1-4440)

QY 16 Lysatgphaeapheargprolysprowaprottyrcysglnalalytyrthrphcys 35
DB 563 AAGAAAGATTCATTGAGCTTACCGAATCTTCTCTGTGCACACGCCACGACAGT 622
QY 36 Prothrngly-----SerProtlepro 42
DB 623 CTTCAAGAAAATGCTTCCCTGTCTCCAGAAAGAGCCTCCATTAACTCTGTTCT 682
QY 43 Valmetgluglyasp-----Aspaspillegluval-----Pheargleu 55
DB 603 CTTAAACGCTGATTTGCTGCGATGATGACAGGTGACAGACAGCAAGAAAGAGAG 742
QY 56 Glnalaprovaltrpgluphelytyrglyaspheuleuglyhisleuylsilemethis 75
DB 743 GAGTCAGGGCAATGAGTGTGAAGATGATGCTTT----- 778
QY 76 Aspalalleglypheargserthrleuthrglylysasenttyrthmetglutrytrglu 95
DB 779 -----GATGAGCCCATGAGAGTTGAAGAG 832
QY 96 LeupheglnleuglyAsnCyserthrpherohisleuargproglumetAspalaprophe 115
DB 803 GTG-----GACCTGAGAGCT--ATGGCTCCCAAGGCT 832
QY 116 TrpCysaenglnlyAlaAlaCysephhegluglyleaspaervalhisleuylsilemethis 135
DB 833 TGGGCAAAAGAGAGT-----GAGCCAGCAAGAGAAAGTGAACAGAGCG 877
QY 136 Asn-----GlyThreuleuValGlnValAlaThrIleserGlyAsnMetPheasnln 152
DB 878 GATTTCGGAAAGGAGCGGTG-----TCTACTTAAGAGATTTCTCCCGAT 925
QY 153 Metalaalystrp---VallyseGlnaspaengluhtrrglyleTytyrGluThrTrpasn 171
DB 926 GTCTCTGTGGACATTCATCAAGAGGTAGTACAGTTTCTCAGTCAAGAAAGTTCA 935
QY 172 Vallysalaserpro-----Gluylsrglyagluhtr----- 132
DB 986 GTGATTCACAGTCACCTCCATTTGTTAAAGAGGAGATGAGGACAAAGTATTCACATT 1345
QY 183 ---TrrpheaapserlyrAsp-----CysseerlysphevalleuargThrPheasn 138
DB 1046 TATTGGTTGGATGCTTATAGAGATGATACAAACCAAGAGGTGAGTATTTCTGTTGGG 1105
QY 199 LysleualagluPhegllyaglu-----Pheylsasnle 210

```

```

DB 1106 AAAGTTGGATTTGATCAGCCGAGACCCATGAGCTGTGTGTGCTGTAATAATATC 1165
QY 211 GlnthraenttyrthrargilepheleutyserglugluProthrTyrluenglyAsnGlu 230
DB 1166 GAGCCAGACG-----CTTTAC-----TTCCTTCCCGTGA 1195
QY 231 ThrservalphegylprothrglyAsnlysthr---leuglyleuAlalleysargphe 249
DB 1196 ATGAAAATTTGATCTTAATAATCGGGGAAAGAACGAGACTCCATATTTCATGAAAGATGT 1255
QY 250 TyrrtyrProphelyProhisleuprothriysglupheuleuSerleuengluile 269
DB 1256 TATGAGATTTGATGAGAAAATACCAAAATATTAATATATAGAC----- 1303
QY 270 PheapalavalilevalhislyseGlnpheTyrluPheTyrluPheGluTyrrphe 289
DB 1304 TTCAAGCTCAAGCCAGTGAAGAAAGAAC-----TATGCTTTGAGATACCTGAT 1351
QY 290 LeuprometylyrpheroPhelileylsilehtrtyrglu-----Gluilepro 305
DB 1352 GTTCAGAAAATCTGAGTACTTGAGAGTTAAATCTCGGCTGAAGTCCACAGCTTCT 1411
QY 306 LeuproileargAsnlyserThrleuser 314
DB 1412 CAAGATTGAAGAGAAACTTTTCT 1438

RESULT 15
US-09-156-842-1
; Sequence 1, Application US/09156842A
; Patent No. 6103473
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Screening
; FILE REFERENCE: STDU-03485
; CURRENT APPLICATION NUMBER: US/09/156,842A
; EARLIER FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-156-842-1

Alignment Scores:
Pred. No.: 1.37      Length: 4440
Score: 88.50      Matches: 71
Percent Similarity: 36.39%      Conservative: 56
Best Local Similarity: 20.34%      Mismatches: 115
Query Match: 5.03%      Indels: 107
DB: 3      Gaps: 20

US-10-010-050a-2_COPY_31_346 (1-316) x US-09-156-842-1 (1-4440)

QY 16 Lysatgphaeapheargprolysprowaprottyrcysglnalalytyrthrphcys 35
DB 563 AAGAAAGATTCATTGAGCTTACCGAATCTTCTGTGTGCACACGCCACGACAGT 622
QY 36 Prothrngly-----SerProtlepro 42
DB 623 CTTCAAGAAAATGCTTCCCTGTCTCCAGAAAGAGCCTCCATTAACTCTGTTCT 682
QY 43 Valmetgluglyasp-----Aspaspillegluval-----Pheargleu 55
DB 603 CTTAAACGCTGATTTGCTGCGATGATGACAGGTGACAGACAGCAAGAAAGAGAG 742
QY 56 Glnalaprovaltrpgluphelytyrglyaspheuleuglyhisleuylsilemethis 75
DB 743 GAGTCAGGGCAATGAGTGTGAAGATGATGCTTT----- 778

```

```
QY 76 Aspa1a1leglyPheArgSerThrLeuThrGlyLysAsnTyThrMetGluTrpTyGlu 35
Db 779 -----GATGAGCCCATGGAAGTTGAAGAG 302
QY 96 LeuPheGlnLeuGlyAsnCysThrPheProHISLeuArgProGluMetAspAlaProPhe 115
Db 803 GTG-----GACCTGAGCCT--ATGGCTGCCAAGGCT 332
QY 116 TrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyLeuAspAspValHISTrpLysGlu 135
Db 833 TCGGACAAAGAGAGT-----GAGCCAGACAGAGAGTGAACAAAGAGCG 377
QY 136 Asn-----GlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 152
Db 878 GATTCGTGGAAAGGACCGTG-----TCCCTACTAGAGAGTTTCTCCCGGAT 325
QY 153 MetAlaLysTrp---ValLysGlnAspAsnGluThrGlyIleTyTyGluThrTrpAsn 171
Db 926 GTCCTCTGTGGACATTCATGACAAAGGTGATGACAGTTTCTCAGTCAAGAAATTCAA 385
QY 172 ValLysAlaSerPro-----GluLysGlyAlaGluThr----- 182
Db 986 GTGATTCACAGTCCACTCCCTCCCATTTGTTAAAGGGCAGATGAGAAACAAGTATTCACCTT 1045
QY 183 ---TrpPheAspSerTyArgPhe-----CysSerLysPheValLeuArgThrPheAsn 198
Db 1046 TATGTGTGGATGCTGTATGAGATCAGTACAAACCAACAGAGTGTGTAATTCGTTGGG 1105
QY 199 LysLeuAlaGluPheGlyAlaGlu-----PheLysAsnIle 210
Db 1106 AAAGTTTGATTCATGATGACGCCGAGACCATGTGAGCTGTGTGCATGTGAAAATATATC 1165
QY 211 GluThrAsnTyThrArgGlyPheLeuTySerGlyGluProThrTyLeuGlyAsnGlu 230
Db 1166 GAGCGAAG-----CTTAC-----TTCTTCCCGGTGA 1195
QY 231 ThrSerValPheGlyProHnGlyAsnLysThr---LeuGlyLeuAlaIleLysArgPhe 249
Db 1196 ATGAAATTCATTAATACGGGAAAGAAACAGAACTCCAAATTCATGAAGAGATGT 1255
QY 250 TyTyTyProPheLysProHISLeuProThrLysGluPheLeuLeuSerLeuGlnIle 269
Db 1256 TATGAGGAATTTGATGAGAAATATGCAACAAATATTAATATGAAAG----- 1303
QY 270 PheAspAlaValIleValHISLysGlnPheTyLeuPheTyAsnPheGluTyTrpPhe 289
Db 1304 TTCAGTCTAAGCCAGTGAAGAAAGAAC-----TATGCTTTGAGATACCTGAT 1351
QY 290 LeuProMetLysPheProPheIleLysIleThrTyGlu-----GluIlePro 305
Db 1352 GTTCCAGAAAAATCTGAGTGAAGTTAAATACCGGCTGAATGCCACAGCTTCT 411
QY 306 LeuProIleArgAsnLysThrLeuSer 314
Db 1412 CAAGATTGAAGGAAACCTTTTCT 1438
```

Search completed: April 25, 2004, 05:51:01  
Job time : 140.894 secs



XX Sheppard PO, Gilbertson DG;  
 XX WPI; 1999-142930/12.  
 DR P-BSDb; AAM92967.  
 XX  
 PT New secreted polypeptide, zsig46, and its fragments, related fusion  
 PT proteins - used for diagnosis and treatment of thyroid disorders or  
 PT diseases involving genes on chromosome 13.  
 XX  
 PS Claim 27; Page 88-90; 101pp; English.

CC This invention describes the isolation of a novel human secreted protein,  
 CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in  
 CC the thyroid. This product can be used to study secretion of proteins from  
 CC cells and also to treat or prevent deficient expression of zsig46, which  
 CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'  
 CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that  
 CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's  
 CC disease, neuronal ceroid-lipofuchnosis, Wilson disease and Reiger  
 CC syndrome). Antibodies and other binding proteins, are used as immunoassay  
 CC reagents to detect zsig46 or cells expressing it, e.g. for assessing  
 CC thyroid function to produce anti-idiotypic antibodies, for affinity  
 CC purification of zsig46, to screen expression libraries, to neutralise  
 CC zsig46 activity and to deliver toxins, radioisotopes etc. for  
 CC therapeutic or diagnostic purposes. Agonists of the product can be used  
 CC to promote growth, differentiation and proliferation of specific cell  
 CC types, e.g. for treating (extra)thyroid diseases or as additive to cell  
 CC cultures  
 XX  
 XX Sequence 1466 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;

Alignment Scores:  
 Score: 1.82e-188 Length: 1486  
 Score: 1758.00 Matches: 316  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AAX02855 (1-1486)

QY 1 ArgValSerGlyIleProSerArgHISITPProValProTyrIleAspGlyPhe 20  
 DB 137 CGGGTCTCGGGACATCCCTCCGCCGCCACTGCGCGGTGCTTCAAGGCGTTGACTTC 196  
 QY 21 ArgProIleProAspProTyrCysGlnAlaIleTyrThrPheCysProthrgIleSerPro 40  
 DB 197 CGTCCAAACCTGATCTTATTCTCAAGCTAAGTAACTTTCTGTCAACTGGCTACCT 256  
 QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgIleGlnAlaProValTyr 60  
 DB 257 ATCCCGATTATGGAGGGTATGATGACATTGAAGTTTTCATTACAAACCCCAAGATCG 316  
 QY 61 GluPheIleTyrGlyAspIleuGluGlyIleuGlyIleuGlyIleuGlyIleuGly 80  
 DB 317 GAATTTAAATATGAGAACCTCTCGGACACTTGAATAATATGATGATCATTCATTGATTG 376  
 QY 81 ArgSerThrLeuThrGlyIleAspAspTyrThrMetGluTyrGluIleuPheGluIleuGly 100  
 DB 377 AGAAGTACATTAACCTGCAAGAACTACCAATGAAATGATGAACTTTTCCAACTGGC 436  
 QY 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGly 120  
 DB 437 AACTGTACATTTCCCATCTCCGACCTGAATGATGATGCCCTTTCTGGTATATCAAGCC 496  
 QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValIleSTPryGluAsnGlyThrIleuVal 140  
 DB 497 GCTGCTGCTTTTITGAGGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 556  
 QY 141 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValIleGlnAsp 160  
 DB 557 CAAGTACCAATATATCAGGAACATGTTTCAACCAAAATGCGAAAGTGGTGAACAGAC 616

QY 161 AsnGluThrGlyIleTyrTyrGluThrThrAsnValIleAspProGluIleGlyAla 180  
 DB 617 AATGAACACAGAAATTTATATGAGACATGAAATGTAAACCCAGCCAGAAAGGGGCA 676  
 QY 181 GluThrTrpPheAspSerTyrAspCysSerIlePheValIleuArgThrPheAsnIleu 200  
 DB 677 GAGACATGTTTATATCTTCACTGACCTGTTCCAAATTTGTATAGAGACCTTTAAACAATG 736  
 QY 201 AlaGluPheGlyAlaGluPheIleuAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220  
 DB 737 GCTGAATTTGACACAGATTCAAGACATAGAACCACTATACAGAAATATTTCTTTAC 796  
 QY 221 SerGluGluProThrTyrIleuGlyIleuGluThrSerValPheGlyProThrGlyAsnIle 240  
 DB 797 AGTGAAGAACCTTACTATCTGGGAAATGAAACATCTGTTTGGCCCAAGCAAG 856  
 QY 241 ThrLeuGlyIleuAlaIleIleAspGlyTyrTyrProPheIleProHisLeuProThrIle 260  
 DB 857 ACTCTTGTTTACCTCAATTAAGATTTTATACCTTCAACCACTTGGCAACTTAA 916  
 QY 261 GluPheIleuSerIleuGluIlePheAspAlaValIleValHisIleGlnPheTyr 280  
 DB 917 GAATTTCTGTGAGTCTCTTGCAAATTTTGAATGAGATGATGATGACAAACAGTTCTAT 976  
 QY 281 LeuPheTyrAspPheGluTyrTyrPheLeuProMetIlePhePheProPheIleIleThr 300  
 DB 977 TGTATTATATATTTGAAATATGATGATTTTACCATGATAATCCCTTTTATTAATAAACA 1036  
 QY 301 TYRGlulIleProLeuProIleArgAsnIleThrLeuSerGlyIleu 316  
 DB 1037 TATGAAGAAATCCCTTACTTACTATCAAGAAACAAACACTCTGCTGTTA 1084  
 RESULT 2  
 AAZ24826  
 ID AAZ24826 standard; DNA; 1751 BP.  
 XX  
 AC AAZ24826;  
 XX  
 DT 02-DEC-1999 (first entry)  
 XX  
 XX Human secreted protein gene 16 clone HMZAD77.  
 DB  
 XX Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumor; neurodegenerative disorder; leukemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
 KW inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 OS Homo sapiens.  
 XX  
 XX WO9947540-A1.  
 PN  
 XX 23-SEP-1999.  
 PD  
 XX 18-MAR-1999; 99WO-US005804.  
 PR  
 XX 19-MAR-1998; 98US-0078563P.  
 PR 19-MAR-1998; 98US-0078566P.  
 PR 19-MAR-1998; 98US-0078573P.  
 PR 19-MAR-1998; 98US-0078574P.  
 PR 19-MAR-1998; 98US-0078575P.  
 PR 19-MAR-1998; 98US-0078577P.  
 PR 19-MAR-1998; 98US-0078578P.  
 PR 19-MAR-1998; 98US-0078579P.  
 PR 19-MAR-1998; 98US-0078581P.  
 PR 01-APR-1998; 98US-0080312P.  
 PR 01-APR-1998; 98US-0080313P.  
 PR 01-APR-1998; 98US-0080314P.  
 PR 01-APR-1998; 98US-0080314P.  
 XX



(HUMA-) HUMAN GENOME SCI INC.

WPI: 1999-562050/47.  
P-PsDB: AAY1323.

new isolated human genes, useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.

Claim 1; Page 308; 484pp; English.

This sequence represents a nucleic acid molecule which encodes a secreted human protein. The gene number, and the clone it is derived from, are detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin  $\mu$  portion (e.g. AA224802) for increasing the stability of the fused protein as compared to the human protein only. The invention relates to 95 novel genes and their fragments (nucleic acid sequences: AA224811-224907; amino acid sequences AA41308-741404) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 95 polynucleotides, based on which tissues they are most highly expressed in (see AA224811 for described uses).

Sequence 1751 BP; 477 A; 355 C; 380 G; 523 T; 0 U; 16 Other;

**Alignment Scores:**

Prod. No.:	4.03e-187	Length:	175
Score:	1747.00	Matches:	314
Percent Similarity:	99.37%	Conservative:	0
Best Local Similarity:	99.37%	Mismatches:	2
Query Match:	99.37%	Indels:	0
JB:	2	Gaps:	0

JS-10-010-050A-2\_COPY\_31\_346 (1-316) X AA224826 (1-1751)

1y	1	L	ArgValIseGlyIleProSerxArgIstPProValProYrLYrAsxPhaAsPhe	20
1b	139	C	GGGATCGGGGATCCCTCCCGGGCCATCGCGGAGCCCTTACAGCCGTTTGACTTC	158
1y	21	A	ArgPcOLySPROAsPcIOYrCYeGlnAlaLYrYrIthPheCSPrOthGlySerPro	40
1b	199	C	CTCCAAACCGATCCTTATTGTCAACCTAAGTATACCTTCTGTCCACTGGCTCACT	258
1y	41	I	IlePcOVAIleMetGluGlyAspAspAspIleGluValPheAlaGluGlnAlaProValITP	60
1b	259	A	ATCCCAAGTTAAGAGAGGATGATGATGACATTGAAGTTTTCATTACAGCCCCAGTATGG	318
1y	61	G	luPheLYrTYGlyAsPLeuLeuGlyYHIsleuLYsIleMetHisAspAlaIleGlyPhe	80
1b	319	G	AATTAAATATGAGAGACCTCCTCGGACACTTGGAATAATTAATCATGATGCCATTGGATTC	378
1y	81	A	ArgSerThrIleuThrGlyLYrAsnTYrIthMetGluITPTrYGlyIleuPheGlnLeuGly	100
1b	379	A	GAAGTACATTAACTGCGCAAGACATACACATGAAATGGTATGAACTTTTCCAACTGGC	438
1y	101	A	AsnCYsThrPhePcOHIsleuArgProGluMetSerAspIaPcProPheTrpCYsAsnGlnGly	120
1b	439	A	ACTGTACATTTCCCACTCCGACCTCGAAGTGAATGATGCCCTTTCTGTGTAAACAAGGC	498
1y	121	A	AlaAlaCYsPhePheGluGlyIleAspAspValHisTrpLYrGluAsnGlyIthLeuVal	140
1b	499	G	CTCGCTGCTCTTTTGGAGGAAATTGATGATGCTTCACTGGAAGGAATATGGGACATTAGTT	558
1y	141	G	luValAlaThrIleIleSerGlyAsnMetPheAsnGluMetAlaLYrTrpValLYrGlnAsp	160
1b	559	C	AACTAGCACTAATATACGAAACATATTTCAACCAATGGCAAGTGGTGGAAACAGGAC	618

QY	161	ANGGUTHGGLYLTETVYRGIUHTHTRPASNVALYALSERPROGLUYSGLYALA	180
Db	619	AATGAAACAGGAATTTATTATAGACATCGAATGTAAAACCCAGCCAGAAAAGGGGCA	678
QY	181	GIUHTHTRTPHASPSERTYRASPQYSSERTLYSPHEVALLEUAQTRHPHEANLYSLEU	200
Db	679	GAGACATGGTTGATTCCCTACACAGCTGTTCCAAATTGTGTAAAGACCTTTAACAGTTG	738
QY	201	MAAGIUPHEGLYALAGIUPHETYSANILEGIUHTHTRASTTYRTHRGLIEPHELEUTYR	220
Db	739	GCTGAATTTGGAGCAGAGTTCAAGAACATVAGAACCACTATAACAGAAATATTTCTTTAC	798
QY	221	SERGIYGIUPROTHRYTRILEUGLYASNGIUHTHRSERVALPHEGLYPROTHRGLYASNLYS	240
Db	799	AGTGAAGAAACCTACTTATCTGSGAAATGAACATCTGTTTGGGCCAACAGSAAACAG	858
QY	241	THRIEUGLYLEUALAILLELYSAIRGPHETYYTRYRPROPHETYSPROHISLEUPROTHRLYS	260
Db	859	ACTCTTGTTTAPGCCATTAATAAAGATTTTATTATCCCTTCAAAACCATTTGGCCACTTAA	918
QY	261	GIUPHELEULESERLEULEUGNILEPHEAPALAVALLLEVALHISLYSGINPHETYR	280
Db	919	GAATTTCTGTGAGAGCTCTTGCAATTTTGTGATGACGATGTGCAACAACAGTTCAT	978
QY	281	LEUPHETRYRASPHEGLUTYRTRIPHELEUPROMETLYSPHEPROPHETILEYSLIETAR	300
Db	979	TGTGTTTATAATTTTGAATATATGGTITTTTACCTATGAATATTCCTTTTATTAATAATPACA	1033
QY	301	TYRGIUGIUILEPROLEUPROILEARGAANLYSTHRLSEUSERYLEU	316
Db	1039	TATGAAGAATCCCTTTACTATCGAAGAACAAAACACTCTCTGGTTTA	1086
RESULT 3			
AAK94829	ID	AAK94829 standard; cDNA; 2120 BP.	
AC	AAK94829;		
XX	06-NOV-2001	(first entry)	
DT	06-NOV-2001	(first entry)	
DE	Human full-length cDNA, SEQ ID NO: 3977.		
XX	Human, full length cDNA; cDNA synthesis; oligo-capping; ss.		
OS	Homo sapiens.		
XX	EP1130094-A2.		
PN	05-SEP-2001.		
PD	07-JUL-2000; 2000EP-00114089.		
XX	08-JUL-1999; 99JP-00194486.		
PR	11-JAN-2000; 2000JP-00118774.		
PR	02-MAY-2000; 2000JP-00183765.		
XX	(HELI-) HELIX RES INST.		
XX	Ota T, Nishikawa T, Isogai T, Hayashi K, Iehi S, Kawai Y;		
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;		
PI	WPI: 2001-524255/58.		
DR	P-PSDB; AAM93870.		
XX	830 Primers useful for synthesizing full length cDNA clones and their use		
PT	in genetic manipulation.		
XX	Claim 8; SEQ ID NO 3977; 1380BP + Sequence Listing; English.		
XX	The invention relates to primers for synthesizing full length cDNA		
CC	clones. 830 cDNA molecules encoding a human protein have been isolated		
CC	and nucleotide sequences of 5' - and 3' -ends of the cDNA molecules have		

CC been determined. Primers for synthesizing the full length cDNA are useful  
 CC for clarifying the function of the protein encoded by the cDNA. The full  
 CC length clones were obtained by construction of full length enriched cDNA  
 CC libraries that were synthesized by the oligo-capping method. The primers  
 CC enable the production of the full length cDNA easily without any special  
 CC methods. The present sequence is a full length human cDNA of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in CD-ROM format directly  
 CC from EPO

XX Sequence 2120 BP; 590 A; 409 C; 460 G; 661 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.:	1.16e-186	Length:	2120
Score:	1744.00	Matches:	314
Percent Similarity:	99.37%	Conservative:	0
Best Local Similarity:	99.37%	Mismatches:	2
Query Match:	99.20%	Indels:	0
	4	Gaps:	0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AAK94829 (1-2120)

```

QY 1 ArgValSerGlyLeuProSerArgArgHisTrpProValProThyTyrAspPheAspHe 20
D 142 CGGGTTCGGGCATCCCTCCGGCCGACATGCGCGATGCGCGAAGCGCTTGGACTTC 201
QY 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
D 202 CGTCCAAACCTGATCTTATGTCTCAAGTAAGTATACCTTGTCTCCAACTGGCTACCT 261
QY 41 IleProValMetGlnGluLysAspAspPheIleGluValPheArgLeuGlnAlaProValTrp 60
D 262 ATCCCGATTAAGAGGAGGTATGATGACACTGAACTTTTCGATTCACACCCCGATGATG 321
QY 61 GluPheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPhe 60
D 322 GAATTTAAATATGAGAACCTCCCGGACACTTGAAATTAATGATATGCAATTTGATTC 381
QY 81 ArgSerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 100
D 382 AGAAGTACATTAACCTGCAAGAACATCAACATGATGATGATGATCTTTCCAACTGGC 441
QY 101 AsnCysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120
D 442 AACTGTACATTTCCCATCTCCGACCTGAAATGGAATGCCCTTCTGATGATCAAGGC 501
QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTrpLysGlnLeuGlyThrLeuVal 140
D 502 GCTGCTGCTTTTATGAGGAATGATGATGATGATGATGATGATGATGATGATGATGAT 561
QY 141 GlnValAlaThrTleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp 160
D 562 CAAGTACCACTATATCAGGAACATGTTCAACCAATGCGCAAGTGGTGAACAACGAGC 621
QY 161 AsnGlnThrGlyLysLeuTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAla 180
D 622 AATGAACACGGAATTTATATATGACATGAAATGAAAGCCAGCCCAAAAAAGGGGGA 681
QY 181 GluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 200
D 682 GAGACATGGTTGATCTTCCACGACCTGTTCAAAATTTGTGTAAAGACCTTTAAACAGTTG 741
QY 201 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyr 220
D 742 GCGAATTTGGAGCAAGTTCAGAAACATAGAAACCAATATACAGATATTTCTTTC 801
QY 221 SerGluGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 240
D 802 AGTGGAGAACCTATATCTGGGAATGAAACATCTGTTTGGGCAACAGAAACAAG 861
QY 241 ThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLys 260
D 862 ACTCTTGTTGATGCGCATTAAGATTTTATTAACCTTCAAAACACATTTGCACTAA 921

```

```

QY 261 GluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyr 280
D 922 GAATTTCTGTTAGCTCTCTGGAATTTTATGATGACGATGTGACAAACAGTTCTAT 981
QY 281 LeuPheTyrAsnPheGlnTyrTrpPheLeuProMetLysPheProPheIleLysIleThr 300
D 982 TTGTTTATTAATTTTGAATATATGTTTATCTATGAATATCCCTTTTATTAATTAACA 1041
QY 301 TyrGluGlnIleProLeuProIleArgAsnLysTrpLeuSerGlyLeu 316
D 1042 TATGAAGAAATCCCTTATCTATGAGAAACAAACACATCTCTGTTTA 1089

```

```

RESULT 4
AAK02866
ID AAK02866 standard; DNA; 1038 BP.
AC AAK02866;
XX
XX 14-MAY-1999 (first entry)
D 14-MAY-1999 (first entry)
DE Human degenerate zsig46 DNA.
KW Secreted protein; zsig46; human; chromosome 13; thyroid; disease;
KW hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;
KW Hirschsprung's disease; neuronal ceroid-lipofuchinosis; Wilson disease;
KW Reiger syndrome; immunosassay; detection; anti-idiotypic antibody;
KW therapy; diagnostic; ss.
OS Homo sapiens.
XX
XX WO9905275-A1.
XX
XX 04-FEB-1999.
PD
XX
XX 24-JUL-1998; 98MO-US015431.
PF
XX
XX 24-JUL-1997; 97US-0053613P.
PR
XX
XX (ZYMO ) ZYMOGENETICS INC.
PA
XX
XX Sheppard PO, Gilbertson DG;
PI
XX
XX WPI; 1999-142930/12.
DR
XX
XX New secreted polypeptide, zsig46, and its fragments, related fusion
PT proteins - used for diagnosis and treatment of thyroid disorders or
PT diseases involving genes on chromosome 13.
XX
XX Claim 31; Page 94-95; 101pp; English.
PS
XX
XX This invention describes the isolation of a novel human secreted protein,
XX zsig46 encoded by a gene on chromosome 13 which is mainly expressed in
XX the thyroid. This product can be used to study secretion of proteins from
XX cells and also to treat or prevent deficient expression of zsig46, which
XX may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
XX disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
XX involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
XX disease, neuronal ceroid-lipofuchinosis, Wilson disease and Reiger
XX syndrome). Antibodies and other binding proteins, are used as immunoassay
XX reagents to detect zsig46 or cells expressing it, e.g. for assessing
XX thyroid function to produce anti-idiotypic antibodies, for affinity
XX purification of zsig46, to screen expression libraries, to neutralise
XX zsig46 activity, and to deliver toxins, radioisotopes etc. for
XX therapeutic or diagnostic purposes. Agonists of the product can be used
XX to promote growth, differentiation and proliferation of specific cell
XX types, e.g. for treating (extra)thyroid diseases or as additive to cell
XX cultures
SQ Sequence 1038 BP; 189 A; 106 C; 179 G; 166 T; 0 U; 398 Other;
Alignment Scores:
Pred. No.: 3.3e-161 Length: 1038

```

Score: 1515.00 Matches: 267  
 Percent Similarity: 85.03% Conservative: 0  
 Best Local Similarity: 85.03% Mismatches: 47  
 Query Match: 86.18% Indels: 0  
 DB: 2 Gaps: 0

US-10-010-050a-2\_copy\_31\_346 (1-316) x AAK02866 (1-1036)

```

QY 2 ValSerGlyIleProSerArgArgHisTrpProValProTyrIysArgPheAspPheArg 21
DB 94 GTNMSGNGNATHCCWMSNMGNMGNCAATGGCCNGTNCNTYAAAMGNTTGAATYTMGN 153
QY 22 ProLysProAspProTyrCysGlnAlaIysTyrThrPheCysProThrGlySerProIle 41
DB 154 CCMNAARCCNGAYCCNTATATGTCARCCNAARTRACNTTGTGTCACACGWSNCCNATH 213
QY 42 ProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrpGlu 61
DB 214 CCNGTATAGARGGAGAYGAYATHGARGTNTTGMNTNCARCCNGTWTGGAR 273
QY 62 PheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArg 81
DB 274 TTYAARTAYGNGAYNTYNTNGNCAYTNAARATGATGAYGAYGCAATGGATYTMGN 313
QY 82 SerThrLeuThrGlyLysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsn 101
DB 334 MSNACNTYNAACNGNAARAAATTAACNATGARGTGTGARGRYNTTYTCARYTMGNARAY 393
QY 102 CysThrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAla 121
DB 394 TGYACNTTYCCNCAATYTMGNCCNGARATGAGYGCNCNTTGTGTGAYAACARGNGCN 453
QY 122 AlaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuValGln 141
DB 454 GCNTGYTYYTYYGARGNATGAGAYGATNCAYTGARGARAAAYGGACVYTMGTTCAR 523
QY 142 ValAlaThrIleSerGlyLysMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsn 161
DB 514 GTNCCNACNATHMSGNMAATATGTTTAAATCAATGGCNAARATGGCTNNAARCAARGAAY 573
QY 162 GluThrGlyIleTyrTyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGlu 181
DB 574 GARACNGNATHTYTAYGARAACNTGGAAYGTNAARCGCWSNCCNGARAAARGNGCNAR 633
QY 182 ThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeuAla 201
DB 634 ACNTGATTYGAYMSNTAYGATYGSNNAARTYGTNTNMGNACNTTAAAYAAAYTMGCN 693
QY 202 GluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLeuTyrSer 221
DB 694 GARTTYGGNGCNGARTTYAARAAYATHGARGCNAAYTRACNMGNAHTHTTYNTATATWSN 753
QY 222 GlyGluProThrTyrTyrLeuGlyAsnGlnThrSerValPheGlyProThrGlyAsnLysThr 241
DB 754 GGNARCCNACNTYTYNNGNAAAGARACWSNCTNTTGGCCNACNGGAAAYAAARACN 813
QY 242 LeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProThrLysGlu 261
DB 814 YTMGNATYNGNATHAARMGNTTATATATCCNTTYAACCCNCAAYTMCCNACNAAAGAR 873
QY 262 PheLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnPheTyrLeu 281
DB 874 TTYTNTYTNMSNYNTYNTNCARATHTTYGAGCNGTNTHTGNCAYARCARITTYATAYTN 933
QY 282 PheTyrAsnPheGluTyrTrpPheLeuProMetLysPheProPheIleLysIleThrTyr 301
DB 934 TTYTAAAYATYTGARTYTGTYTNTYNTNCNATGARGARTTYCCNTTYATHAARATACNTAY 953
QY 302 GluGluIleProLeuProIleArgAsnLysThrLeuSerGly 315
DB 994 GARCARATHCCNTYNTCCNATHMGNAAYAAARACNTTMSNGSN 1035

```

```

AAK93424
ID AAK93424 standard; cDNA; 697 BP.
XX
AC AAK93424;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA clone representative sequence, SEQ ID NO: 1884.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Example 11; SEQ ID NO 1884; 1380bp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesizing the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesized by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence was used as the representative sequence
CC from a human clone which was used in homology searches to identify the
CC clone. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in CD-ROM format directly from
CC EPO
XX
SQ Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;

```

Alignment Scores:

Pred. No.:	Length:	Score:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1	9e-91	896.00	170	92.51%	90.91%	50.97%	2

Percent Similarity: 92.51%  
 Best Local Similarity: 90.91%  
 Query Match: 50.97%

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AAK93424 (1-697)

```

QY 1 ArgValSerGlyIleProSerArgArgHisTrpProValProTyrIysArgPheAspPhe 20
DB 142 CGGATCTCGGGCAATCCCTCCGCGCACTGGCCGGTGCCTGCAAGCGCTTGACTTC 201
QY 21 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPheCysProThrGlySerPro 40
DB 202 CGTCCAAAACCTATCTTATTCATCAAGCTAACTATACCTTCTGTCCAACTGGCTCACT 261
QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgLeuGlnAlaProValTrp 60
DB 262 ATCCCAATTATGAGAGGTGATGATGACACGAGAGTTCGATTACCAAGCCCAAGATATGG 321

```

QY 61 GluPheIysTyrGlyAspIleuLeuGlyIleuIleuMetHisAspAlaIleGlyPhe 80  
 DB 322 GAATTTAAATATGAGACCTCTCGGACACCTGAAATATCATGATGCGATTGGAATTC 381  
 QY 81 ArgSerThrLeuThrGlyIleAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 100  
 DB 382 AGAAGTACATTTAACTGCGAAGAACTACACATGAAATGATGAACTTTCCAACTTGCC 441  
 QY 101 AsnCyThrPheProHisIleuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120  
 DB 442 AACTGTACATTTCCCATCTCCGACCTGAAATGAAATGAGCCCTTTCTGGTAAATCAAGGC 501  
 QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTrpIlyGluAsnGlyThrIleu-Val 140  
 DB 502 GGTGCTGCTTTTGGAGGGAATGATGATGCTTCACTGGAAGAAAGGGAAGGCACTTAAGT 561  
 QY 140 IgluValAlaThrIleSer-GlyAsnMetPheAsnGluMetAla-IlyTrpValIlyGln 159  
 DB 562 TCAAGTACATTTAACTGCGAAGAACTGTTCAACCAATGCAAAAGTGGGTGAACAA 621  
 QY 160 ---AspAsnGluThrGlyIleTyrTyrGlu-ThrTrpAsnValIlyAlaSerProGluIly 178  
 DB 622 GGACATATGAAACAAGAAATTTATTATGAGAACATGGAATTTTAAAGCCANCCANAA 681  
 QY 178 S---GlyAlaGlu 181  
 DB 682 AAANGGGGCAAAA 694  
 RESULT 6  
 AAK92158  
 ID AAK92158 standard; cDNA; 697 BP.  
 AC AAK92158;  
 DT 06-NOV-2001 (first entry)  
 XX Human cDNA 5'-end sequence, SEQ ID NO: 618.  
 DE Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX EP133094-A2.  
 PN 05-SEP-2001.  
 PD 07-JUL-2000; 2000BP-00114089.  
 PF 08-UTL-1999; 99TP-00194486.  
 PR 11-JAN-2000; 2000BP-00118774.  
 PR 02-MAY-2000; 2000BP-00183765.  
 XX (HELI-) HELIX RES INST.  
 PA Ota T, Nishikawa T, Isegai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 DR MPI; 2001-524255/58.  
 XX 830 Primers useful for synthesizing full length cDNA clones and their use  
 PT in genetic manipulation.  
 XX Claim 2; SEQ ID NO 618; 1380BP + Sequence Listing; English.  
 XX The invention relates to primers for synthesizing full length cDNA  
 CC clones. 830 cDNA molecules encoding a human protein have been isolated  
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have  
 CC been determined. Primers for synthesizing the full length cDNA are useful  
 CC for clarifying the function of the protein encoded by the cDNA. The full  
 CC length clones were obtained by construction of full length enriched cDNA  
 CC libraries that were synthesized by the oligo-capping method. The primers  
 CC enable the production of the full length cDNA easily without any special  
 CC methods. The present sequence is the nucleotide sequence of the 5'-end of

CC a cDNA provided in the invention. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in CD-  
 CC ROM format directly from EPO  
 XX  
 SQ Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;  
 Alignment Scores:  
 Pred. No.: 1.9e-91 Length: 697  
 Score: 896.00 Matches: 170  
 Percent Similarity: 92.51% Conservative: 3  
 Best Local Similarity: 90.91% Mismatches: 8  
 Query Match: 50.97% Indels: 6  
 DB: 4 Gaps: 2  
 US-10-010-050A-2\_copy\_31\_346 (1-316) x AAK92158 (1-697)  
 QY 1 ArgValSerGlyIleProSerArgHisTrpProValProTyrIlyAsnGlyPheAspPhe 20  
 DB 142 CGGGTCTCGGCAATCCCTCCCGGCGCACTGAGCGCGGCTCGCAAGCGCTTGAACTTC 201  
 QY 21 ArgProIlyProAspProTyrCysGlnAlaIlyTyrThrPheCysProThrGlySerPro 40  
 DB 202 GCTCCAAACCTGATCCTTATTGTCAGCTAAGTACTTCTGTCCAACTGCGTCACT 261  
 QY 41 IleProValMetGluGlyAspAspAspIleGluValPheArgIleuGlnAlaProValTrp 60  
 DB 262 ATCCCACTTATGAGGGGATGATGACACTGAAGTTTTCATTACAAGCCCAAGTATGG 321  
 QY 61 GluPheIysTyrGlyAspIleuLeuGlyIleuIleuMetHisAspAlaIleGlyPhe 80  
 DB 322 GAATTTAAATATGAGACCTCTCGGACACCTGAAATATCATGATGCGATTGGAATTC 381  
 QY 81 ArgSerThrLeuThrGlyIleAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGly 100  
 DB 382 AGAAGTACATTTAACTGCGAAGAACTACACATGAAATGATGAACTTTCCAACTTGCC 441  
 QY 101 AsnCyThrPheProHisIleuArgProGluMetAspAlaProPheTrpCysAsnGlnGly 120  
 DB 442 AACTGTACATTTCCCATCTCCGACCTGAAATGAAATGAGCCCTTTCTGGTAAATCAAGGC 501  
 QY 121 AlaAlaCysPhePheGlnGlyIleAspAspValHisTrpIlyGluAsnGlyThrIleu-Val 140  
 DB 502 GGTGCTGCTTTTGGAGGGAATGATGATGCTTCACTGGAAGAAAGGGAAGGCACTTAAGT 561  
 QY 140 IgluValAlaThrIleSer-GlyAsnMetPheAsnGluMetAla-IlyTrpValIlyGln 159  
 DB 562 TCAAGTACATTTAACTGCGAAGAACTGTTCAACCAATGCAAAAGTGGGTGAACAA 621  
 QY 160 ---AspAsnGluThrGlyIleTyrTyrGlu-ThrTrpAsnValIlyAlaSerProGluIly 178  
 DB 622 GGACATATGAAACAAGAAATTTATTATGAGAACATGGAATTTTAAAGCCANCCANAA 681  
 QY 178 S---GlyAlaGlu 181  
 DB 682 AAANGGGGCAAAA 694  
 RESULT 7  
 AAI25096  
 ID AAI25096 standard; DNA; 494 BP.  
 AC AAI25096;  
 DT 12-OCT-2001 (first entry)  
 XX Probe #15029 for gene expression analysis in human cervical cell sample.  
 DE Probe; human; microarray; gene expression; cervical epithelial cell;  
 KW cervical cancer; ss.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200157278-A2.

```
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX
XX Claim 25; SEQ ID NO 15029; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging of
XX diseases of the cervix, notably cervical cancer. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.57e-90 Length: 494
Score: 886.00 Matches: 163
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.40% Indels: 0
DB: Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x AAI25096 (1-494)
QY 154 AlAlysrTPVallysgInAspAsngluThrglylletYrYrgluThrTPAsnVallys 173
DB 3 GCAAAAGTGGTGAACAGGACATGAAACAGAAATTATATGAGACATGGAATGTAAAA 62
QY 174 AlAserProglulYgLyAlGluThrTPheAspserTYrAspCYsSerlyspheVal 193
DB 63 GCCAGCCCAAGAAAGGGGGCAGACACATGCTTATCTTCGACCTGTCACAAATTTGTG 122
QY 194 LeuarGrThrPheAnlyslLeuAlAGluPheglYAlAGluPheYsAnllGluThrAsn 213
DB 123 TTAAGGACCTTAAACAGCTTGGCTGAATTTGAGAGAGATTCAAGAACATAGAAACCAAC 182
QY 214 TyrThArgllePheLeuTYrSerGlygluPProThrTYrLeuGlyAsngluThrSerVal 233
DB 183 TATACAGAAATATTTCTTTACAGTGAACCTACTTATCTGGGAATGAAACATCTGTT 242
QY 234 PheglYProThrGlyAsnlyslThleuGlyLeuAllelylsArGpheyTYrTYrProPhe 253
DB 243 TTTGGGGCCAAACAGAAACAGACTCTGTTAGCCATAAAAAATTTTATTAACCCCTTC 302
QY 254 LysProhlsleuPProThrlysgInPheleuSerleuengnlllePheAspAlaVal 273
DB 303 AAACCAATTGGCCAACTAAAGAAATTCGTGTGAGCTCTTGCAAAATTTTGATGACGTG 362
QY 274 lLevalhlslysgInpheyTYrleuPheTYrAsnPhelYrTYrPheleuProMeLys 293
DB 363 ATGTGCAAAACAGTCTCTATTGTTTATTAATTTTGAATATTTGTTTATACCATGAAA 422
```

```
QY 294 PheProhelleYslleThTYrGluGluileProleuProlleAsnlyslThleu 313
DB 423 TTCCCTTTATTAATAATACATATGAAAGAAATCCCTTACCTATCAGAAACAAACACTTC 482
QY 314 SerGlylyl 316
DB 483 TCTGGTTTA 491

RESULT 8
ABA70786
ID ABA70786 standard; DNA; 494 BP.
XX
XX ABA70786;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #19091.
XX
XX Human, foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX
XX Claim 4; SEQ ID NO 19091; 633bp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.57e-90 Length: 494
Score: 886.00 Matches: 163
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.40% Indels: 0
DB: Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x ABA70786 (1-494)
QY 154 AlAlysrTPVallysgInAspAsngluThrglylletYrYrgluThrTPAsnVallys 173
DB 3 GCAAAAGTGGTGAACAGGACATGAAACAGAAATTATATGAGACATGGAATGTAAAA 62
```

```

QY 174 AAlSerProGluLySGlYAlaGluThrTrpPheAspSerTyrAspCysSerIysPheVal 193
DB 63 GCCAGCCCGAAGAAAGGGGCGAGACATGGTTGATTCTTACGACTGTTCCAAATTGTG 122
QY 194 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 213
DB 123 TTAAAGACCTTTAACAAGATTGGCTGAATTTGGAGCAAGATTCAGAAACATAGAAACCAAC 182
QY 214 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233
DB 183 TATACAGAGATATTTCTTTACAGTGAAGAACCTACTATCTGGAAATGAAACATCTGTT 242
QY 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProbe 253
DB 243 TTGGGCGCAACAGAAACAGACTCTTGTTAGCATATAAAGATTATTAACCCCTTC 302
QY 254 LysProHisLeuProThrLysGluPheLeuLysSerLeuGluIlePheAspAlaVal 273
DB 303 AAACCACTTTGCCAATTAAGAAATTTCTGTGAGCTCTTGCAAATTTTATGCACTG 362
QY 274 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMetLys 293
DB 363 ATTGTGCAACAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGAAA 422
QY 294 PheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 313
DB 423 TTCCCTTTATTAATAATACATATGAGAAATCCCTTACTATCAGAAACAAACACTC 482
QY 314 SerGlyLeu 316
DB 483 TCTGTTTA 491

```

## RESULT 9

AA150961  
ID AA150961 standard; DNA; 494 BP.

XX AC AA150961;

DT 17-OCT-2001 (first entry)

DE Probe #19647 used to measure gene expression in human placenta sample.

KM Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.

OS Homo sapiens.

PN MO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US000663.

PR 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-0063236P.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-48897/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX PT gene expression in human placenta.

PS Claim 25; SEQ ID NO 19647; 654bp; English.

XX CC The present invention relates to single exon nucleic acid probes (SENPs).

CC The present sequence is one such probe. The probes are useful for  
CC producing a microarray for predicting, measuring and displaying gene  
CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders

XX SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

## Alignment Scores:

Pred. No.:	1-57e-90	Length:	494
Score:	886.00	Matches:	163
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	50.40%	Indels:	0
DB:	4	Gaps:	0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AA150961 (1-494)

```

QY 154 AAlaYsTPValIlySGlnAspAsnGluThrGlyIleTyrTyrGluThrTrpAsnVallys 173
DB 3 GCAAAGTGGTGAAACAGACATGAAACAGAAATTTATATGAGACATGGAATGTAA 62
QY 174 AAlSerProGluLySGlYAlaGluThrTrpPheAspSerTyrAspCysSerIysPheVal 193
DB 63 GCCAGCCCGAAGAAAGGGGCGAGACATGGTTGATTCTTACGACTGTTCCAAATTGTG 122
QY 194 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 213
DB 123 TTAAAGACCTTTAACAAGATTGGCTGAATTTGGAGCAAGATTCAGAAACATAGAAACCAAC 182
QY 214 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233
DB 183 TATACAGAGATATTTCTTTACAGTGAAGAACCTTACTTGTGGAAATGAAACATCTGTT 242
QY 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProbe 253
DB 243 TTGGGCGCAACAGAAACAGACTCTTGTTTATATTTGAATTTGTTTACCTATGAAA 302
QY 254 LysProHisLeuProThrLysGluPheLeuLysSerLeuGluIlePheAspAlaVal 273
DB 303 AAACCACTTTGCCAATTAAGAAATTTCTGTGAGCTCTTGCAAATTTTATGCACTG 362
QY 274 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMetLys 293
DB 363 ATTGTGCAACAACAGTCTATTTGTTTATATTTGAATTTGTTTACCTATGAAA 422
QY 294 PheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 313
DB 423 TTCCCTTTATTAATAATACATATGAGAAATCCCTTACTATCAGAAACAAACACTC 482
QY 314 SerGlyLeu 316
DB 483 TCTGTTTA 491

```

## RESULT 10

ABA37272  
ID ABA37272 standard; DNA; 494 BP.

AC ABA37272;

DT 23-JAN-2002 (first entry)

DE Probe #15738 for gene expression analysis in human heart cell sample.

KM Human; gene expression; heart; microarray; vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

OS Homo sapiens.

PN MO200157274-A2.

PD 09-AUG-2001.

```

PF 30-JAN-2001; 2001WO-US000666.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 4; SEQ ID NO 15738; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease. Note: The sequence data for this patent did not
XX form part of the printed specification, but was obtained in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pcr_sequences
XX
XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1.57e-90 Length: 494
XX Score: 886.00 Matches: 163
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 50.40% Indels: 0
XX DB: Gaps: 0
XX
XX US-10-010-050A-2_COPY_31_346 (1-316) x ABA37272 (1-494)
XX
XX 154 AAlaLyfTrpValLySGlnAspAsnGluThrGlyLeuTyrTrpGluThrTrpAsnValLys 173
XX 3 GCMAAGTGGGTGAAACAGACGACATGAACAGAAATTATTATGACATGATGATGAAA 62
XX
XX 174 AlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSerLysPheVal 193
XX 63 GCCAGCCCGAAGAAAGGGGGGAGAGACATGGTTGATTCCTACACATGTTCCAAATTTG 122
XX
XX 194 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLysGluThrAsn 213
XX 123 TTAAAGACCTTTAAACAAGTTGGCTGAATTTGGAGGAGAGTTCAAGAAACATAGAAACCAAC 182
XX
XX 214 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrIleuGlyAsnGluThrSerVal 233
XX 183 TAAACAAAGAAATATTTCTTACAGTGAAGAACCTACTATCTGGAAATGAAACATCTGTT 242
XX
XX 234 PheGlyProThrGlyAsnLysThrLeuGlyLeuValIleLysArgPheTyrTrpProPhe 253
XX 243 TTGGGGCCACAGCAAGAAACAGACTCTTGTTTGCCATTAAGAAAGATTATTAATCCCTTC 302
XX
XX 254 LysProIleLeuProThrLysGluPheLeuSerLeuLeuGlnIlePheAspAlaVal 273
XX 303 AAACCAATTGGCACTAAAGAAATTTCTGTGAGTCTCTGCAATTTTGTATGACAGTG 362
XX
XX 274 IleValIleLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMetLys 293
XX 363 ATTGTGCAAAACAGTTCTAATTTGTTTAAATTTGAATATGATTTTAACTATGAAA 422
XX

```

```

XX 294 PheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 313
XX 423 TTCCCTTTATTTAAATTAACATATGAAAGAAATCCCTTACTACTACAGAAACAAACACTC 482
XX
XX 314 SerGlyLeu 316
XX 483 TCTGGTTTA 491
XX
XX RESULT 11
XX AAK44992
XX ID AAK44992 standard; DNA; 494 BP.
XX
XX AC AAK44992;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 19549.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000666.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48890/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 19549; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention
XX
XX Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1.57e-90 Length: 494
XX Score: 886.00 Matches: 163
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 50.40% Indels: 0
XX DB: Gaps: 0
XX
XX US-10-010-050A-2_COPY_31_346 (1-316) x AAK44992 (1-494)
XX
XX 154 AAlaLyfTrpValLySGlnAspAsnGluThrGlyLeuTyrTrpGluThrTrpAsnValLys 173
XX 3 GCMAAGTGGGTGAAACAGACGACATGAACAGAAATTATTATGACATGATGATGAAA 62
XX
XX 174 AlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSerLysPheVal 193
XX

```



DB 63 GCCAGCCGAGAAAAGGGGAGAGACATGGTTGATTCCTACGACTGTTCCAAATTGTG 122  
QY 134 LeuArgThrPheAsnIleuAlaGluPheGlyAlaGluPheIleuAsnIleGluThrAsn 213  
DB 123 TTAAGGACCTTTAAACAAGTGGCTGAATTTGAGAGAGAGTCAAGAACATGAACACAC 182  
QY 214 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233  
DB 183 TATACAGAAATTTCTTTACAGTGAAGACCTACTATCTGGGAAATGAACATCTGTT 242  
QY 234 PheGlyProThrGlyAsnIleuThrLeuGlyLeuAlaIleuAsnGlyPheTyrTyrProPhe 253  
DB 243 TTTGGGCCACAGAAACAGACTCTGGTTAGCATTAAGAAAGATTTATATACCCCTTC 302  
QY 254 LysProHisIleuProThrTyrLeuGluPheLeuSerIleuGluIlePheAsnAlaVal 273  
DB 303 AAACCAATTTGCCAACAATAAGAAATTCGTGAGTCTCTGCACAAATTTGATCAGTG 362  
QY 274 ILevalHisIleGlnPheTyrLeuPheTyrAsnPheGlyTyrTyrPheLeuProMetLys 293  
DB 363 ATTGAGCAAAACAGTCTATTTGTTATATAATTTGATATATGGTTTATACCTATGAA 422  
QY 294 PheProPheIleuIleTyrIleuGluIleProLeuProIleuArgAsnIleuThrLeu 313  
DB 423 TTCCCTTTTATTAATAATACATATGAGAAATCCCTTACCTATCAGAAACAAACACTC 482  
QY 314 SerGlyLeu 316  
DB 483 TCTGTTTA 491  
RESULT 12  
AAK19040  
ID AAK19040 standard; DNA; 494 BP.  
AC AAK19040;  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 19031.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
KW ss.  
XX  
OS Homo sapiens.  
XX  
PN WC000157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001MO-US000667.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
XX  
PR 26-MAY-2000; 2000US-0207456P.  
XX  
PR 30-JUN-2000; 2000US-00608408.  
XX  
PR 03-AUG-2000; 2000US-00632366.  
XX  
PR 21-SEP-2000; 2000US-0234687P.  
XX  
PR 27-SEP-2000; 2000US-0236359P.  
XX  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX  
PT brains.  
XX  
PS Example 4; SEQ ID NO 19031; 650bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention  
XX  
SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 1.57e-90 Length: 494  
Score: 886.00 Matches: 163  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 50.40% Indels: 0  
DB: 4 Gaps: 0  
US-10-010-050a-2\_copy\_31\_346 (1-316) x AAK19040 (1-494)  
QY 154 AlaIleuTyrValIleGlnAsnIleuAlaGluThrTyrPheAsnIleGluThrAsn 173  
DB 3 GCAAGTGGGTGAAACAGGACATGAACACAGAAATTTATATGAGACATGAAATGTA 62  
QY 174 AlaSerProGluIleuGlyAlaGluThrTyrPheAsnIleGluThrAsn 193  
DB 63 GCCAGCCGAGAAAAGGGGAGAGACATGGTTGATTCCTACGACTGTTCCAAATTGTG 122  
QY 194 LeuArgThrPheAsnIleuAlaGluPheGlyAlaGluPheIleuAsnIleGluThrAsn 213  
DB 123 TTAAGGACCTTTAAACAAGTGGCTGAATTTGAGAGAGAGTCAAGAACATGAACACAC 182  
QY 214 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 233  
DB 183 TATACAGAAATTTCTTTACAGTGAAGACCTACTATCTGGGAAATGAACATCTGTT 242  
QY 234 PheGlyProThrGlyAsnIleuThrLeuGlyLeuAlaIleuAsnGlyPheTyrTyrProPhe 253  
DB 243 TTTGGGCCACAGAAACAGACTCTGGTTAGCATTAAGAAAGATTTATATACCCCTTC 302  
QY 254 LysProHisIleuProThrTyrLeuGluPheLeuSerIleuGluIlePheAsnAlaVal 273  
DB 303 AAACCAATTTGCCAACAATAAGAAATTCGTGAGTCTCTGCACAAATTTGATCAGTG 362  
QY 274 ILevalHisIleGlnPheTyrLeuPheTyrAsnPheGlyTyrTyrPheLeuProMetLys 293  
DB 363 ATTGAGCAAAACAGTCTATTTGTTATATAATTTGATATATGGTTTATACCTATGAA 422  
QY 294 PheProPheIleuIleTyrIleuGluIleProLeuProIleuArgAsnIleuThrLeu 313  
DB 423 TTCCCTTTTATTAATAATACATATGAGAAATCCCTTACCTATCAGAAACAAACACTC 482  
QY 314 SerGlyLeu 316  
DB 483 TCTGTTTA 491  
RESULT 13  
ABS44658  
ID ABS44658 standard; DNA; 494 BP.  
AC ABS44658;  
DT 25-FEB-2003 (first entry)  
XX  
DE Human liver single exon probe, SEQ ID NO 19648.  
XX  
XX  
KW Human; single exon nucleic acid probe; liver; cirrhosis;  
KW hyperlipoproteinemia; hyperlipidaemia; hypercholesterolaemia;  
KW coronary heart disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WC000157273-A2.  
XX  
PD 09-AUG-2001.

```
KX 30-JAN-2001; 2001MO-US000664.
ZF
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
(KX)
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
KX WPI; 2001-488898/53.
KX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
PS Claim 4; SEQ ID NO 19648; 658bp; English.
KX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridizes at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS1005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
KX
SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1.57e-90 Length: 494
Score: 886.00 Matches: 163
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 50.40% Indels: 0
DB: Gaps: 0

US-10-010-050A-2_COPY_31_346 (1-316) x ABS44658 (1-494)
QY 154 AAlayrtpvlllyvglnaspaangluThrcglylletYrYrgluThrtPAsnVallye 173
DB 3 GCAAGTGGGTGAACAGGACATGAAACAGAAATTATATGAGACATGGAATGTAATA 62
QY 174 AAlaserpogluylvglyalagluThrtPheaspserYrAapCyseSeLyPheVal 193
DB 63 GCCAGCCCAAGAAAAGGGGCGACAGCATGTGTTGATCTCTACGATGCTTCCAAATTTGTC 122
QY 194 LeuArygThrPheasnlyleuAlagluPheglYAlagluPheylasnllegluThrasn 213
DB 123 TTAAGGACCTTTAACAAGTTGGTGAATTGGACAGAGTTCAAGAACATAGAAACCAAC 182
QY 214 TyrThraagliepHeleuTyrserygluPProhrtYrleuuglyaaangluThserval 233
DB 183 TATACAAATAATTTCTTTACAGTGAAGAACCTTACTTATCTGGGAAATGAAACATCTGTT 242
QY 234 PheglYProThrglyAsnlyThrluuglyleuAlalyleYsAlyPheTyrtYrProPhe 253
DB 243 TTTGGGCCAACAAGAAACAAGACTTGGTTTACCCARAAAAAATTTATACCCCTTC 302
QY 254 LysProhileuPProThryleugluPheleuLeuSerleuLeuGlnliePheAspAlaVal 273
DB 303 AAACCACTTGGCAACTTAAGAAATTTCTGTGAGTCTCTTCAAAATTTTGATGACGTG 362
```

```
QY 274 lIeValHlelysglnPheTyrlleuPheTyAsnPhelgluTyrtPheleuProweLys 293
DB 363 ATTGTGCAACAAGAGTCTATATTTGTTTATATTTGAATTTGATTTTCTTACCTAGAAA 422
QY 294 PheProPheIlelyslleThrTyrgluGluileProleuProIleArgAsnlysthrLeu 313
DB 423 TTCCCTTTATTAATAATACATATGAAGAAATCCCTTACCTATCGAAACAAACGCTC 482
QY 314 SerGlyLeu 316
DB 483 TCTGGTTTA 491

RESULT 14
ABS19237
ID ABS19237 standard; DNA; 494 BP.
XX
AC ABS19237;
XX
AC 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon ORF from lung SEQ ID No 19228.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; IID;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemoideriosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001MO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DB WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 4; SEQ ID NO 19228; 634bp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12817 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC ; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
```

CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probe/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe open reading frame of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

# Alignment Scores:

Pred. No.: 1.57e-90 Length: 494  
Score: 886.00 Matches: 163  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 50.40% Indels: 0  
Gaps: 0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x A8S19237 (1-494)

QY 154 AAlaYsTrpValYsGlnAspAsnGluThrGly11eYrYrGluThrTrpAsnValYs :.73  
Db 3 GCAAGGCGGTGAACAGCAATGAAACAGAAATTTATTATGACATGAAATGTAATAA 62  
QY 174 AlAserProGluYsGlyAlaGluThrTrpPheAspSerYrAspCySerYsPheVal :.93  
Db 63 GCCAGCCCAAGAAAGGGGGGAGAGACATGTTGATTCCTAACGACCTTCCAAATTTGTG :.22  
QY 194 LeuAarghrPheAsnYsLeuAlaGluPheGlyAlaGluPheYsAsn11eGluThrAsn :.23  
Db 123 TTAAAGACCTTTAAACAAGTTGGCTGAATTTGGAGCAAGTTTCAAGACATAGAAACCAAC :.82  
QY 214 TyrThrArg11ePheLeuYrSerGlyGluProThrYrLeuGlyAsnGluThrSerVal :.23  
Db 183 TATACAGAAATATTTCTTTACAGTGAAGAACTTACTATCTGGGAAATGAAACATCTGTT :.242  
QY 234 PheGlyProThrGlyAsnYsThrLeuGlyLeuAla11eYsAsnGlyYrYrProPhe :.253  
Db 243 TTTGGGCAACAGCAAGAAACAAGCTCTGTGTTACCAATAAAATTTTATACCCCTTC :.302  
QY 254 YsPheProHsLeuProThrYsGluPheLeuYsSerLeuGlu11ePheAspAlaVal :.273  
Db 303 AAACCAAGCTTTGCCAAGTAAAGAAATTTCTGTGAGCTCTTGCAGAAATTTGATGCGTG :.362  
QY 274 11eValHsYsGlnPheYrLeuPheYrAsnPheGlyYrYrPheLeuProMetYs :.293  
Db 363 ATTGTGCAACAAGCTTCTATTTGTTTAAATTTGAAATTTGGTTTATACCTATGAAA :.422  
QY 294 PheProHelleYs11eThrYrGluGlu11eProLeuPro11eAsnYsThrLeu :.313  
Db 423 TTTCCCTTTTATAAATAACATATGAAGAATCCCTTACTATCAAGAAACAAACACCTC :.482  
QY 314 SerGlyLeu 316  
Db 483 TCTGCTTTA 491

RESULT 15  
AAV88347  
ID AAV88347 standard; cDNA; 506 BP.

XX AAV88347;

XX 12-FEB-1999 (first entry)

XX EST clone GB814.

XX Expressed sequence tag; secreted protein; haematopoiesis regulator;  
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytic;  
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.

XX Homo sapiens.

XX MO9845437-A2.

XX 15-OCT-1998.

XX 10-APR-1998; 98WO-US006956.

XX 10-APR-1997; 97US-00837312.

XX (GENE) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;  
PI Spaulding V, Agostino MJ;

XX WPI; 1999-070078/06.

XX New polynucleotides encoding human secreted proteins - derived from e.g.  
PT human blood, kidney, foetal lung, placenta, testes, brain, ovary,  
PT pituitary, retina and colon cDNA libraries.

XX Claim 1, Page 366; 641pp; English.

XX The present sequence represents an expressed sequence tag (EST), and is a  
CC polynucleotide of the invention. The polynucleotides of the invention are  
CC all secreted EST sequences isolated from a variety of human tissue  
CC sources. The EST sequences and proteins encoded by them are predicted to  
CC have useful biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals, although no supporting data is given. Suggested activities  
CC include nutritional activity, immune stimulating or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
CC activity. The EST sequences are also stated to be useful for gene therapy

XX Sequence 506 BP; 171 A; 86 C; 119 G; 130 T; 0 U; 0 Other;

# Alignment Scores:

Pred. No.: 1.56e-72 Length: 506  
Score: 727.00 Matches: 133  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 41.35% Indels: 0  
Gaps: 0

US-10-010-050a-2\_COPY\_31\_346 (1-316) x AAV88347 (1-506)

QY 121 AAlaCyPhePheGluGly11eAspAsnValHsTrpYsGluAsnGlyThrLeuVal :.140  
Db 99 GCTGCTGCTTTTGAAGGAATTGATGATGTTCACTGGAAGAAATAGGACATTAGTT :.158  
QY 141 GlnValAlaThr11eSerGlyAsnMetPheAsnGlnMetAlaYsTrpValYsGlnAsp :.160  
Db 159 CAAGTAGCAACTATATCAGGAAACATGTTCAACCAATGCGAAAGGGTGAACAGGAC :.218  
QY 161 AsnGluThrGlyLeuYrYrGluThrTrpAsnValYsAlaSerProGluYsGlyAla :.180

```
Db 219 AATGAAACAGAAATTTATTATGAGACATGAAATGTAATAAGCCAGCCAGAAAGGGGCA 278
Qy 181 G|u|t|h|r|T|P|p|h|e|S|e|r|T|y|A|s|p|C|y|S|e|r|y|S|p|h|e|V|a|l|l|e|u|A|g|T|h|r|P|h|e|A|s|n|y|S|l|e|u 200
Db 279 GAGACATGCTTGAATTCCTACGACTGTTCCAAATTTGTGTTAAGGACCTTTAACAGTTG 318
Qy 201 A|A|g|u|P|h|e|G|y|A|A|g|u|P|h|e|y|S|n|l|l|e|G|u|T|h|r|A|s|n|y|T|h|r|A|g|l|l|e|P|h|e|u|T|y|r 220
Db 339 GCTGAATTTGAGCAGAGATTCAAGAACATAGAAACCACTATACAGAAATATTTCTTTAC 398
Qy 221 S|e|r|G|y|G|u|P|r|o|T|h|r|T|y|L|e|u|G|y|A|s|n|G|u|T|h|r|S|e|r|V|a|l|P|h|e|G|y|P|r|o|T|h|r|G|y|A|s|n|y|S 240
Db 399 AGTGGAGAACCTTACTTATCTGGGAAATGAAACATCTGTTTGGGCCAACAGAAACAG 458
Qy 241 T|h|r|L|e|u|G|y|L|e|u|A|a|l|l|e|y|S|A|g|P|h|e|T|y|T|y|r|P|r|o|P|h|e 253
Db 459 A|C|T|C|T|G|T|T|A|G|C|C|A|T|A|A|A|A|G|A|T|T|T|A|T|T|A|C|C|C|C|T|T|C 497
```

Search completed: April 24, 2004, 23:26:28  
Job time : 589.283 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:02:32 ; Search time 380.792 Seconds

(without alignments)  
2352.639 Million cell updates/sec

Title: US-10-010-050A-2\_COPY\_1\_30  
Perfect score: 169  
Sequence: 1 MRGAGAGAGGASCMALALMLAVPGMS 30

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODB=frame+ p2n.model -DEV=xlh  
-O=/cg2.1/USPT0.spool/US10010050/runat.22042004.113204.27556/app.query.fasta\_1.1372  
-DB=EST -OFMT=fasta -SUFFIX=first -MINMATCH=0.1 -IOOPT=0 -IOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human0.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORE=PCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10010050 @CGN 1 1.6283 @runat.22042004.113204.27556 -MCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOC  
-BEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hcc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	* Match	Query Length	DB ID	Description
1	169	100.0	565	9	AUT280382
2	169	100.0	607	12	BM849262
3	169	100.0	623	13	BQ417448
4	169	100.0	676	12	BG820113
5	169	100.0	787	12	B1223533
6	169	100.0	847	13	BQ719074
7	169	100.0	869	12	B1090566
8	169	100.0	923	13	BX331615
9	169	100.0	1098	10	BF982158
10	169	100.0	1201	9	AL546472
11	169	100.0	1222	13	BQ897986
12	161	95.3	924	14	CA488543
13	156	92.3	1077	13	BX342662
14	142	84.0	589	10	BF977311
15	135	79.9	555	14	CA397574
16	119	70.4	273	9	AA300650
17	73	43.2	383	14	CB808908
18	73	43.2	424	14	CB795968
19	73	43.2	427	14	CB548129
20	73	43.2	1338	13	BU792679
21	72.5	42.9	127	28	CC378302
22	72.5	42.9	497	29	CG375783
23	72.5	42.9	587	10	BG077284
24	72.5	42.9	713	28	CC30780
25	72.5	42.9	730	29	CG108116
26	72.5	42.9	732	29	CG108118
27	72.5	42.9	825	29	CG453851
28	72.5	42.9	830	29	CG230524
29	72.5	42.9	834	29	CG230534
30	72.5	42.9	860	29	CG610458
31	72.5	42.9	2940	11	AK028566
32	72	42.6	651	29	AG105984
33	72	42.6	742	14	CF876196
34	72	42.6	855	14	CB904966
35	72	42.6	1355	10	BE880806
36	71.5	42.3	1025	10	BE567958
37	71	42.0	569	29	CE827309
38	71	42.0	632	13	BX844586
39	71	42.0	641	14	CB215456
40	70.5	41.7	840	28	AQ898634
41	70.5	41.7	401	9	AA108220
42	70.5	41.7	427	9	AA062254
43	70.5	41.7	494	12	BG101391
44	70.5	41.7	547	12	BG101262
45	70.5	41.7	591	14	CF106239

## ALIGNMENTS

RESULT 1  
AUT280382  
LOCUS AUT280382 NIESE2 Homo sapiens cDNA clone NIESE2001763 5', mRNA  
DEFINITION  
SEQUENCE  
ACCESSION AUT280382  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

565 bp mRNA linear EST 31-JUL-2003  
Sequence.  
AUT280382.1 GI:28299609  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 565)

COMMENT

Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-339, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409

ORGANISM	REFERENCE
Human sapiens	1 (baes 1 to 623)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Euteria; Primates; Carnivora; Canidae; Canis	
Melomys, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K.,	
Hammer, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,	
Hillier, L., Merra, M., Page, D., Wylie, T., Martin, J., Blissett, A.,	
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,	
Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R.,	
Williams, T., Jackson, Y., and Bowers, Y.	
Endocrine Pancreas Consortium	
Unpublished (2000)	
Other ESTs: ik8b05.xl	
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue	
Endocrine Pancreas Consortium	







```

Db      52 ATGGGCGGGGCGGGGCGGGCTCGGGAGCGCGTTCCTGCTGCTGGGCCCTGCGCTG 111
Oy      21 LeuTIPLeuAlaValAlaProGlyTTPSer 30
Db      112 CTTGGCTCGCGGTGTTCGGGGCTGTGTC 141

RESULT 8
LOCUS   BX331615
DEFINITION BX331615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
          CDNA clone CS0DB009YC01 5-PRIME, mRNA sequence.
ACCESSION BX331615
VERSION   BX331615
KEYWORDS  BX331615.1 GI:30310073
SOURCE    EST.
          Homo sapiens (human)
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 923)
          Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
          Full-length cDNA libraries and normalization
          Unpublished (2001)
COMMENT   Contact: Genoscope
          BP 191 91006 Evry cedex - France
          Genoscope - Centre National de Sequencage
          Email: secrete@genoscope.cns.fr, Web : www.genoscope.cns.fr
          Library was constructed by Life Technologies, a division of
          Invitrogen. This sequence belongs to sequence cluster 7238.f For
          more information about this cluster, see
          http://www.genoscope.cns.fr/
          cgi-bin/cluster.cgi?seq=CS0DB009AB01QPLcluster=7238.f. Contact :
          Feng Liang Email : fliang@lifetech.com URL : Corporation 1600
          http://fulllength.invitrogen.com/Invitrogen 1600
          Parady Avenue Genoscope sequence ID : CS0DB009AB01QPL.

FEATURES
    source
        1..923
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="CS0DB009YC01"
        /tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
        /clone_id="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
        /note="First strand cDNA was primed with a NotI-oligo (dt)
        primer. Five prime end enriched, double-strand cDNA was
        digested with Not I and cloned into the Not I and EcoR V
        sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Alignment Scores:
Pred. No.:      2,06e-07      Length:      923
Score:          169.00      Matches:      30
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             13          Gaps:          0

US-10-010-050a-2_copy_1_30 (1-30) x BX331615 (1-923)
Oy      1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTPCySTIPAlaLeuAlaLeu 20
Db      118 ATGGGCGGGGCGGGGCGGGCTCGGGAGACCGCTTCTGCTGTGGGCCCTGGCGCTG 177
Oy      21 LeuTIPLeuAlaValAlaProGlyTTPSer 30
Db      178 CTTGGCTCGCGGTGTTCGGGGCTGTGTC 207

RESULT 9
LOCUS   BF982158
DEFINITION 602308976r1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4400298 5',
          mRNA sequence.
ACCESSION BF982158

```

```

VERSION   BF982158.1 GI:12384970
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1098)
          NIH-MGC http://mgs.nci.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
JOURNAL   Contact: Robert Strausberg, Ph.D.
          Email: cgabs-remail.nih.gov
          Tissue Procurement: ATCC
          CDNA Library Preparation: Life Technologies, Inc.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LMNL at:
          http://image.jnl.gov
          Plate: LMNL0105 row: 9 column: 19
          High quality sequence stop: 693.

FEATURES
    source
        1..1098
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:4400298"
        /tissue_type="duodenal adenocarcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /clone_id="NIH_MGC_88"
        /note="Organ: small intestine; Vector: pCMV-SPORT6;
        Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
        oligo-dT primed. Average insert size 1,767 kb. Library
        enriched for full-length clones and constructed by Life
        Technologies. Note: this is a NIH_MGC Library."

ORIGIN
Alignment Scores:
Pred. No.:      2,59e-07      Length:      1098
Score:          169.00      Matches:      30
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             10          Gaps:          0

US-10-010-050a-2_copy_1_30 (1-30) x BF982158 (1-1098)
Oy      1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTPCySTIPAlaLeuAlaLeu 20
Db      35 ATGGGCGGGGCGGGGCGGGCTCGGGAGACCGCTTCTGCTGTGGGCCCTGCGCTG 94
Oy      21 LeuTIPLeuAlaValAlaProGlyTTPSer 30
Db      95 CTTGGCTCGCGGTGTTCGGGGCTGTGTC 124

RESULT 10
LOCUS   AL546472
DEFINITION AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
          clone CS0D1030Y01 5-PRIME, mRNA sequence.
ACCESSION AL546472
VERSION   AL546472.2 GI:31268306
KEYWORDS  EST.
          Homo sapiens (human)
          SOURCE
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1201)
          Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
          Full-length cDNA libraries and normalization
          Unpublished (2001)
JOURNAL   On Feb 15, 2001 this sequence version replaced gi:12879620.
COMMENT   Contact: Genoscope

```

Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: sequef@genoscope.cns.fr Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSODI030CE010P1a;cluster=7238.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CSODI030CE010P1.

## FEATURES

source

```
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI030Y01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (d)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

## ORIGIN

Alignment Scores:  
Pred. No.: 2,91e-07 Length: 1201  
Score: 169.00 Matches: 30  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-10-010-050a-2\_COPY\_1\_30 (1-30) x AU546472 (1-1201)

QY 1 MetAgaAGGlyAlaGlyAlaAlaArgGlyArgAlaSerTPCyETPAlaLeuAlaLeu 20  
Db 63 ATGCGGCGGGCGGGCGGGCGGGCTCGGAGACGCGCTCTGCTGCTGCGGCGCTG 122

QY 21 LeuTPLeuAlaValAlaProGlyTPSer 30  
Db 123 CTTTGCTCGCGGTTCGGGTTCGGGTTC 152

## RESULT 11

LOCUS BQ897986 1222 bp mRNA linear EST 16-AUG-2002  
DEFINITION AGENCOURT 8482480 lupski\_dorsal\_root ganglion Homo sapiens CDNA  
clone IMAGE:6185883 5', mRNA sequence.

## ACCESSION

BQ897986

## VERSION

BQ897986

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

Rukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## TITLE

NIH-MGC http://mgc.nci.nih.gov/.

## JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

## COMMENT

Unpublished (1999)

## CONTACT

Contact: Robert Strausberg, Ph.D.

## Tissue

Email: cgapbs-remail.nih.gov

## CDNA Library Preparation

Life Technologies, Inc.

## DNA Sequencing

Agencourt Bioscience Corporation

## Clone distribution

MGC clone distribution information can be

## found through the I.M.A.G.E. Consortium/LNLN at:

## http://image.llnl.gov

## Plate: LHAM13577

row: 9 column: 04

## High quality sequence

start: 277.

## High quality sequence

stop: 277.

## Location/Qualifiers

1..1222

## FEATURES

source

1..1222

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6185883"
/sex="male"
/tissue_type="dorsal root ganglia"
/dev_stage="adult, 36 yr"
/lab_host="MDH10B"
/clone_lib="Lupski dorsal root ganglion"
/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGCTCG-3' and
5'-GACTAGTCTGATCGAGCGGCCCTT(15)-3'. Size selected >
1 kb for average insert length 1.7 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine) and is available through Life
Technologies."
```

## ORIGIN

Alignment Scores:  
Pred. No.: 2,97e-07 Length: 1222  
Score: 169.00 Matches: 30  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 13 Gaps: 0

US-10-010-050a-2\_COPY\_1\_30 (1-30) x BQ897986 (1-1222)

QY 1 MetAgaAGGlyAlaGlyAlaAlaArgGlyArgAlaSerTPCyETPAlaLeuAlaLeu 20  
Db 17 ATGCGGCGGGCGGGCGGGCGGGCTCGGAGACGCGCTCTGCTGCTGCGGCGCTG 76

QY 21 LeuTPLeuAlaValAlaProGlyTPSer 30  
Db 77 CTTTGCTCGCGGTTCGGGTTCGGGTTC 106

## RESULT 12

LOCUS CA488543 924 bp mRNA linear EST 14-NOV-2002  
DEFINITION AGENCOURT 10808884 MABOL Homo sapiens CDNA clone IMAGE:6720309 5',  
mRNA sequence.

## ACCESSION

CA488543

## VERSION

CA488543

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

Rukayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## TITLE

NIH-MGC http://mgc.nci.nih.gov/.

## JOURNAL

National Institutes of Health, Mammalian Gene Collection (MGC)

## COMMENT

Unpublished (1999)

## CONTACT

Contact: Robert Strausberg, Ph.D.

## Tissue

Email: cgapbs-remail.nih.gov

## CDNA Library Preparation

Invitrogen Corp

## DNA Sequencing

Agencourt Bioscience Corporation

## Clone distribution

MGC clone distribution information can be

## found through the I.M.A.G.E. Consortium/LNLN at:

http://image.llnl.gov

## Plate: LHAM14279

row: 3 column: 21

## High quality sequence

stop: 637.

## High quality sequence

stop: 637.

## Location/Qualifiers

1..924

## FEATURES

source

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924

## source

1..924



Best Local Similarity: 96.67%  
 Query Match: 84.02%  
 DB: 10  
 Gaps: 0

US-10-010-050a-2\_COPY\_1\_30 (1-30) x BF977311 (1-555)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20  
 Db 161 ATCGCGCGGGCGCGCGCGCGGT-CGCGGACGCGCTTCCCTGCTGCTGGCCCTGGCGCTG 219  
 QY 21 LeuTrpLeuAlaValAlaProGlyTrpSer 30  
 Db 220 CTTTGGCTCGCGGTGCTCGGCGCTGCTCC 249

RESULT 15  
 CA397574  
 LOCUS  
 DEFINITION

CA397574 555 bp mRNA linear EST 06-NOV-2002  
 cs92h10.y2 Human Retinal pigment epithelium/choroid cDNA  
 (Un-normalized, unamplified): cs Homo sapiens CDNA clone cs92h10  
 5', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS

CA397574.1 GI:24735005  
 EST.  
 SOURCE  
 ORGANISM

Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE  
 AUTHORS  
 TITLE

Wistow, G., Bernstein, S.L., Wyatt, M.K., Farris, R.N., Behal, A.,  
 Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.  
 Expressed sequence tag analysis of human RPE/choroid for the  
 NEIBank Project: Over 6000 non-redundant transcripts, novel genes  
 and splice variants

Mol. Vis. 8 (4), 205-220 (2002)  
 22103460  
 12107410

COMMENT

Contact: Wistow G  
 Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA

Tel: 301 402 3452  
 Fax: 301 496 0076

Email: graeme@helix.nih.gov  
 Plate: 92 row: h column: 10

Seg primer: M13RPI reverse primer (ABI).

Location/Qualifiers

1..555

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="cs92h10"

/tissue\_type="RPE/choroid"

/dev\_stage="Adult"

/lab\_host="EMDH10B"

/clone\_lib="Human Retinal pigment epithelium/choroid cDNA  
 (Un-normalized, unamplified): cs"

/note="Organ: Eye; Vector: PCMVSPORT6; Two different donor  
 eyes (75-80 years old) yielded approximately 600 mg of  
 dissected RPE/choroid tissue. This in turn yielded 340 ug  
 of total RNA and 7 ug of mRNA. A directionally cloned cDNA  
 library in the pCMVSPORT6 vector was constructed at Life  
 Technologies (Rockville, MD; now part of Invitrogen Corp),  
 essentially following the protocols of the Superscript  
 Plasmid System (Invitrogen Corp).  
 <http://www.invitrogen.com/> The library code  
 designation was cs. For this library, cDNA inserts were  
 cloned into the NotI/Mlu sites of the vector. EST  
 analysis was performed on the unamplified library at the  
 NIH Intramural Sequencing Center (NISC)."

ORIGIN

Alignment Scores:  
 Pred. No.:

0.000319

length:

555

Score: 135.00  
 Percent Similarity: 82.76%  
 Best Local Similarity: 82.76%  
 Query Match: 79.88%  
 DB: 14  
 Gaps: 0

US-10-010-050a-2\_COPY\_1\_30 (1-30) x CA397574 (1-555)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20  
 Db 40 ATCGCGCGGGCGCGCGCGCGGT-CGCGGACGCGCTTCCCTGCTGCTGGCCCTGGCGCTG 99  
 QY 21 LeuTrpLeuAlaValAlaProGlyTrp 29  
 Db 100 CTTTGGCTCGCGGTGCTCGGCGCTGCTCC 126

Search completed: April 25, 2004, 05:46:22  
 Job time : 383.792 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:04:07 ; Search time 59.7917 Seconds

(Without alignments)  
2262.168 Million cell updates/sec

Title: US-10-010-050A-2\_COPY\_1\_30  
Perfect score: 169  
Sequence: 1 MRGAGARGARSMCAALMLAVPGWS 30

Scoring table:  
BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=x1h  
-Q=/cgn2\_1/USPRO.spool/US10010050/funat.22042004.113205.27584/app\_query.fasta\_1.1372  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=blomsum62  
-TRANS=human40.cdd -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10010050@cgn2\_1.1.776@funat.22042004.113205.27584  
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WALT -DSPLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	169	100.0	1486	9	US-09-122-383-1	Sequence 1, Appl1
2	169	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appl1
3	169	100.0	1751	13	US-10-653-595-26	Sequence 26, Appl1
4	169	100.0	1751	13	US-09-397-945-26	Sequence 26, Appl1
5	117	69.2	1038	9	US-09-122-383-13	Sequence 13, Appl1
6	117	69.2	1038	14	US-10-010-050A-13	Sequence 13, Appl1
7	71	42.0	580	13	US-10-027-633-24716	Sequence 24716, Appl1
8	71	42.0	580	16	US-10-027-633-24716	Sequence 24716, Appl1
9	71	42.0	1071	13	US-10-282-122A-15190	Sequence 15190, A
10	70.5	41.1	256	9	US-09-960-352-12900	Sequence 12900, A
11	69.5	41.1	403	9	US-09-960-352-12900	Sequence 6089, Ap
12	69.5	41.1	404	9	US-09-960-352-12900	Sequence 11275, A
13	69.5	41.1	414	9	US-09-960-352-12900	Sequence 14095, A
14	68.5	40.5	4853	10	US-09-814-353-21894	Sequence 21894, A
15	68	40.2	4444	16	US-10-395-740-15	Sequence 15, Appl1
16	67.5	39.9	276276	13	US-10-087-754	Sequence 754, App
17	66.5	39.3	3738	15	US-10-029-386-20677	Sequence 20677, A
18	66.5	39.3	3857	13	US-10-087-192-2057	Sequence 2057, Ap
19	66.5	39.3	24990	13	US-10-087-192-2056	Sequence 2056, Ap
20	66	39.1	7155	15	US-10-329-079-14	Sequence 14, Appl1
21	66	39.1	3736	15	US-10-329-079-6	Sequence 6, Appl1
22	65	38.5	468	13	US-10-027-632-181138	Sequence 181138, A
23	65	38.5	468	13	US-10-027-632-181139	Sequence 181139, A
24	65	38.5	468	16	US-10-027-632-181138	Sequence 181138, A
25	65	38.5	468	16	US-10-027-632-181139	Sequence 181139, A
26	64.5	38.2	12951	13	US-10-282-122A-31678	Sequence 31678, A
27	64	37.9	1490	13	US-10-280-576-25	Sequence 25, Appl1
28	64	37.9	1530	13	US-10-282-122A-30126	Sequence 30126, A
29	64	37.9	2000	16	US-10-260-238-1708	Sequence 1708, Ap
30	63.5	37.6	594	9	US-09-950-933A-10	Sequence 10, Appl1
31	63.5	37.6	1704	15	US-10-156-761-1987	Sequence 4987, Ap
32	63.5	37.6	9025608	15	US-10-156-761-1	Sequence 1, Appl1
33	63	37.3	646	13	US-10-027-632-103173	Sequence 103173, A
34	63	37.3	646	16	US-10-027-632-103173	Sequence 103173, A
35	63	37.3	722	15	US-10-029-517-106	Sequence 106, App
36	63	37.3	1257	13	US-10-425-114-5227	Sequence 5227, Ap
37	63	37.3	96599	12	US-09-997-722-262	Sequence 262, App
38	63	37.3	332101	13	US-10-354-247-1	Sequence 1, Appl1
39	63	37.3	322101	15	US-10-060-902-1	Sequence 1, Appl1
40	62.5	37.0	1872	10	US-09-764-891-7060	Sequence 7060, Ap
41	62.5	37.0	1962	13	US-10-411-037-65	Sequence 65, Appl1
42	62.5	37.0	1962	13	US-10-411-026-65	Sequence 299, Appl1
43	62.5	37.0	2155	9	US-09-962-436-299	Sequence 2394, Ap
44	62.5	37.0	2155	9	US-09-880-107-2394	Sequence 1, Appl1
45	62.5	37.0	6200	9	US-09-993-038-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-09-122-383-1  
Sequence 1, Application US/09122383A  
Patent No. US20020042093A1  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
FILE REFERENCE: 97-38  
CURRENT APPLICATION NUMBER: US/09/122,383A  
EARLIER FILING DATE: 1998-07-24  
EARLIER APPLICATION NUMBER: 60/053,613  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1486  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (47)...(1084)

US-09-122-383-1

## Alignment Scores:

Pred. No.:	4,18e-12	Length:	1486
Score:	169.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050A-2\_COPY\_1\_30 (1-30) x US-09-122-383-1 (1-1486)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTTPCySTTPAlaLeuAlaLeu 20

Db 47 ATGCGGGCGGGCGGGCGGGCGGGCTGCGGAGACGGCTTCTGATGCTGGCGGCTGGCGCTG 106

QY 21 LeuTTPLeuAlaValValProGlyTTPSer 30

Db 107 CTTTGCTCGCGGTGCTTCCGGGCTGCTCC 136

## RESULT 2

US-10-010-050A-1

; Sequence 1, Application US/10010050A

; Publication No. US20020173624A1

; GENERAL INFORMATION:

; APPLICANT: Sheppard, Paul O.

; APPLICANT: Gilbertson, Debira G.

; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN

; FILE REFERENCE: 97-38C1

; CURRENT APPLICATION NUMBER: US/10/010,050A

; PRIOR FILING DATE: 2002-03-26

; PRIOR APPLICATION NUMBER: US 09/122,383

; PRIOR FILING DATE: 1998-07-24

; PRIOR APPLICATION NUMBER: US 60/053,613

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 1486

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (47)...(1084)

US-10-010-050A-1

; Alignment Scores:

Pred. No.:	4.18e-12	Length:	1486
Score:	169.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050A-2\_COPY\_1\_30 (1-30) x US-10-010-050A-1 (1-1486)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTTPCySTTPAlaLeuAlaLeu 20

Db 47 ATGCGGGCGGGCGGGCGGGCGGGCTGCGGAGACGGCTTCTGATGCTGGCGGCTGGCGCTG 106

QY 21 LeuTTPLeuAlaValValProGlyTTPSer 30

Db 107 CTTTGCTCGCGGTGCTTCCGGGCTGCTCC 136

## RESULT 3

US-10-653-595-26

; Sequence 26, Application US/10653595

; Publication No. US20040048304A1

; GENERAL INFORMATION:

; APPLICANT: Ruben et. al.

; TITLE OF INVENTION: 95 Human secreted proteins

; FILE REFERENCE: P2027P1C1

; CURRENT APPLICATION NUMBER: US/10/653,595

; CURRENT FILING DATE: 2003-09-03

; PRIOR APPLICATION NUMBER: US 09/397945

; PRIOR FILING DATE: 1999-09-17

; PRIOR APPLICATION NUMBER: PCT/US99/05804

; PRIOR FILING DATE: 1999-03-18

; PRIOR APPLICATION NUMBER: 60/078,566

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,576

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,573

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,574

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/078,579

; PRIOR FILING DATE: 1998-03-19

; PRIOR APPLICATION NUMBER: 60/080,314

; PRIOR FILING DATE: 1998-04-01

; PRIOR APPLICATION NUMBER: 60/080,312

; PRIOR FILING DATE: 1998-04-01

; PRIOR APPLICATION NUMBER: 60/078,578

; PRIOR FILING DATE: 1998-03-19

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 470

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 26

; LENGTH: 1751

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (1520)

; OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (1557)

; OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (1689)

; OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (1735)

; OTHER INFORMATION: n equals a,t,g, or c

; FEATURE:

; NAME/KEY: SITE

; LOCATION: (1741)

; OTHER INFORMATION: n equals a,t,g, or c

; US-10-653-595-26

; Alignment Scores:

Pred. No.:	4.93e-12	Length:	1751
Score:	169.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-10-010-050A-2\_COPY\_1\_30 (1-30) x US-10-653-595-26 (1-1751)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTTPCySTTPAlaLeuAlaLeu 20

Db 49 ATGCGGGCGGGCGGGCGGGCGGGCTGCGGAGACGGCTTCTGATGCTGGCGGCTGGCGCTG 108

QY 21 LeuTTPLeuAlaValValProGlyTTPSer 30

Db 109 CTTTGCTCGCGGTGCTTCCGGGCTGCTCC 138



```

RESULT 4
US-09-397-945-26
; Sequence 26, Application US/09397945
; Publication No. US20030065139A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc. et al.
; TITLE OF INVENTION: 95 Human secreted proteins
; FILE REFERENCE: P2027P1
; CURRENT APPLICATION NUMBER: US/09/397,945
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PCT/US99/05804
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/078,566
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,576
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,573
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,574
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,579
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,314
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,312
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/078,578
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,581
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,577
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,563
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,313
; PRIOR FILING DATE: 1998-04-01
; NUMBER OF SEQ ID NOS: 470
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 26
; LENGTH: 1751
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1520)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1557)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1689)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1729)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1735)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (1741)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-397-945-26

Alignment Scores:
Pred. No.: 4,93e-12 Length: 1751
Score: 169.00 Matches: 30
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 13
US-10-010-050a-2_COPY_1_30 (1-30) x US-09-397-945-26 (1-1751)

```

```

QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
DB 49 ATGGGGCGGGGGCGGGGGCGGGCTTGGGGAGCGCGCTTCTGGTGTGGGCGCTG 108
QY 21 LeuTrpLeuAlaValAlaProGlyTrpSer 30
DB 109 CTTGGCTGGCGGTGTTCGGGGCTGTGC 138

RESULT 5
US-09-122-383-13
; Sequence 13, Application US/09122383A
; Patent No. US20020042093A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38
; CURRENT APPLICATION NUMBER: US/09/122,383A
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 60/053,613
; PRIOR FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Degenerate nucleotide sequence encoding zsig46
; OTHER INFORMATION: polypeptide of SEQ ID NO:2
; NAME/KEY: variation
; LOCATION: (1)...(1038)
; OTHER INFORMATION: n is any nucleotide
US-09-122-383-13

Alignment Scores:
Pred. No.: 9,88e-06 Length: 1038
Score: 117.00 Matches: 20
Percent Similarity: 68.97% Conservative: 0
Best Local Similarity: 68.97% Mismatches: 9
Query Match: 69.23% Indels: 0
Gaps: 0
DB: 9

US-10-010-050a-2_COPY_1_30 (1-30) x US-09-122-383-13 (1-1038)
QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
DB 1 ATGGAGMGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
QY 21 LeuTrpLeuAlaValAlaProGlyTrp 29
DB 61 YTTTGGYTNGCNGTNGTNGCNGGNTGG 87

RESULT 6
US-10-010-050a-13
; Sequence 13, Application US/10010050A
; Publication No. US20020173624A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Gilbertson, Debra G.
; TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
; FILE REFERENCE: 97-38C1
; CURRENT APPLICATION NUMBER: US/10/010,050A
; PRIOR FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: US 09/122,383
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/053,613
; PRIOR FILING DATE: 1997-07-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0

```



```

APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT FILING DATE: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15190
LENGTH: 1071
TYPE: DNA
ORGANISM: Bordetella pertussis
IS-10-282-122A-15190

Alignment Scores:
    Trd. No.:      6.07      Length:      1071
    Score:         71.00     Matches:       12
    Percent Similarity: 53.57%   Conservative:   3
    Best Local Similarity: 42.86%   Mismatches:    13
    Query Match:    42.01%     Indels:        0
    B:              13          Gaps:          0
IS-10-010-050A-2_COPY_1_30 (1-30) x US-10-282-122A-15190 (1-1071)

Y      2 AAGAGGAYALAGYALALAArGYALGAsetTTPcyerTTPAlalaLeuLau 21
      b 251 AGCGCTGGGCAAGCGGTCCGCATTCGCCGCCGACGTGTCTGGGCATGGGCGCTATTG 310
      Y      22 TTPLeuAlaValalPrroGLyTP 29
      b 311 TGCGTTTCCCGGCGCGCATCG 334

ESULT 10
S-09-960-352-12900/c
Sequence 12900, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengping
APPLICANT: Byatt, John C.
APPLICANT: Machialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 12900
LENGTH: 256
TYPE: DNA
```

[illegible]

```

; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11275
; LENGTH: 404
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-LIB34-036-Q1-E1-D8
US-09-960-352-11275

Alignment Scores:
Pred. No.: 3.54 Length: 404
Score: 69.50 Matches: 15
Percent Similarity: 48.48% Conservative: 1
Best Local Similarity: 45.45% Mismatches: 10
Query Match: 41.12% Indels: 7
DB: 9 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-960-352-11275 (1-404)
QY 4 G1YAlaG1YAlaA1aArgG1YArgAlaSerTYPys-----TTP 16
DB 304 GGAGCAGGTGCATACGTCGTCGTCATGTTGCCGTACTTGGGCGTTCTTCATGG 245
QY 17 A1aLeuA1aLeuTTPLeuA1aVala1ProG1YTTP 29
DB 244 GCTCTGGGCGCTGCATGGGCGGCGGCATCAGCATGG 206

RESULT 13
US-09-960-352-14095/C
; Sequence 14095, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Mengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14095
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 60-LIB34-005-Q1-E1-G8
US-09-960-352-14095

Alignment Scores:
Pred. No.: 3.63 Length: 414
Score: 69.50 Matches: 15
Percent Similarity: 48.48% Conservative: 1
Best Local Similarity: 45.45% Mismatches: 10
Query Match: 41.12% Indels: 7
DB: 9 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-960-352-14095 (1-414)
QY 4 G1YAlaG1YAlaA1aArgG1YArgAlaSerTYPys-----TTP 16
DB 402 GGAGCAGGTGCATACGTCGTCGTCATGTTGCCGTACTTGGGCGTTCTTCATGG 343
QY 17 A1aLeuA1aLeuTTPLeuA1aVala1ProG1YTTP 29
DB 342 GCTCTGGGCGCTGCATGGGCGGCGGCATCAGCATGG 304

RESULT 14
US-09-814-353-21894/C
; Sequence 21894, Application US/099814353
; Publication No. US20030165831A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-0063
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21894
; LENGTH: 4853
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3, 4, 4827, 4828, 4829, 4830, 4831, 4832, 4833, 4834,
; LOCATION: 4835, 4836, 4837, 4838, 4839, 4840, 4841, 4842, 4843, 4844,
; LOCATION: 4845, 4846, 4847, 4848, 4849, 4850, 4851, 4852, 4853
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21894

Alignment Scores:
Pred. No.: 56.5 Length: 4853
Score: 68.50 Matches: 14
Percent Similarity: 62.07% Conservative: 4
Best Local Similarity: 48.28% Mismatches: 10
Query Match: 40.53% Indels: 1
DB: 10 Gaps: 1

US-10-010-050A-2_COPY_1_30 (1-30) x US-09-814-353-21894 (1-4853)
QY 1 MetAArgG1YAlaG1YAlaA1aArgG1YArgAlaSerTYPysTTPAlaLeuA1aLeu 20
DB 4590 ATTGGGCGGCTTGTGGTGCGGCC--GGGGGGGTGTGTTTGCTGGCTAGTTTG 4534
QY 21 LeuTTPLeuA1aVala1ProG1YTTP 29
DB 4533 TGCTGGAGTGGCGGTCTTGGCGCGTTGG 4507

RESULT 15
US-10-395-740-15
; Sequence 15, Application US/10395740
; Publication No. US20030215852A1
; GENERAL INFORMATION:
; APPLICANT: WYETH
; APPLICANT: Mark, Robert
; APPLICANT: Wood, Andrew
; APPLICANT: Gulukota, Kamalakhar
; TITLE OF INVENTION: NOVEL PANCREATIN-PABLO PROTEIN INTERACTIONS AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: AM100375PCT
; CURRENT APPLICATION NUMBER: US/10/395,740
; CURRENT FILING DATE: 2003-03-24
; PRIOR APPLICATION NUMBER: US 60/369,244
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: US 60/386,645
; PRIOR FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 15
```

SOFTWARE: PatentIn version 3.2

SEQ ID NO 15

LENGTH: 44442

TYPE: DNA

ORGANISM: Genomic DNA

JS-10-395-740-15

Alignment Scores:

Prod. No.:	595	Length:	44442
Score:	68.00	Matches:	13
Percent Similarity:	53.12%	Conservative:	4
Best Local Similarity:	40.62%	Mismatches:	13
Query Match:	40.24%	Indels:	2
DB:	16	Gaps:	1

JS-10-010-050A-2\_COPY\_1\_30 (1-30) x US-10-395-740-15 (1-44442)

Y 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20

yb 27846 ATGCAAGAGAGGAAAGAGGAGCCCTTGAGCCACGCGAGGTTGGGCTGGGGCAGAGCCGCC 27905

Y 21 LeuTrpLeuAlaValVal-----ProGlyTrpSer 30

yb 27906 CTCTGAGAGCAACTGCAAGGTCCAGGTTGGCG 27941

Search completed: April 25, 2004, 06:14:58  
Job time : 68.7917 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## 4 protein - nucleic search, using frame\_plus\_p2n model

on: April 24, 2004, 23:02:52 ; Search time 9.95833 Seconds

(without alignment)  
1671.819 Alignment cell updates/sec

file: US-10-010-050A-2\_COPY\_1\_30

sequence: 1 MRGAGARGRASWCALALMLAVPGMS 30

scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

searched: 682709 seqs, 277475446 residues

total number of hits satisfying chosen parameters: 1365418

minimum DB seq length: 0  
maximum DB seq length: 2000000000

hit-processing: Minimum Match 0%

Maximum Match 100%  
listing first 45 summaries

command line parameters:

CODE=frame\_p2n.model -DEV=x1h  
p=/cg2\_1/USPTO/spool/US10010050/runat.22042004.113205.27569/app.query.fasta\_1.1372  
JB=Issued Patents NA -OEMT=fastbp -SUFFIX=rmi -MINMATCH=0.1 -LOOFCI=0  
OOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
IST=45 -DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
MODE=LOCAL -OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
JSE=US10010050/cg2\_1.140/runat.22042004.113205.27569 -NCPU=6 -ICPU=3  
JMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOG  
JBTIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
JGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

itabase: Issued Patents NA:\*

1: /cg2\_6/pdata/2/ina/5A.COMB.seq:\*  
2: /cg2\_6/pdata/2/ina/5B.COMB.seq:\*  
3: /cg2\_6/pdata/2/ina/6A.COMB.seq:\*  
4: /cg2\_6/pdata/2/ina/6B.COMB.seq:\*  
5: /cg2\_6/pdata/2/ina/PCTUS.COMB.seq:\*  
6: /cg2\_6/pdata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	64	37.9	438	4	US-09-252-991A-9484	Sequence 9484, Ap
2	64	37.9	594	4	US-09-252-991A-7647	Sequence 7647, Ap
3	64	37.9	1533	4	US-09-252-991A-9378	Sequence 9378, Ap
4	64	37.9	2043	4	US-09-252-991A-9459	Sequence 9459, Ap
5	64	37.9	2274	4	US-09-252-991A-9347	Sequence 9347, Ap
6	63.5	37.6	2316	4	US-09-252-991A-13835	Sequence 13835, A
7	63.5	37.6	2811	4	US-09-252-991A-13619	Sequence 13619, A
8	63	37.3	700	1	US-08-037-579A-4	Sequence 4, Appli
9	63	37.3	700	1	US-08-601-184-4	Sequence 4, Appli
10	62.5	37.0	2067	2	US-08-713-928B-8	Sequence 8, Appli
11	62.5	37.0	2155	3	US-09-191-171-4	Sequence 4, Appli
12	62.5	37.0	2155	3	US-09-385-707-4	Sequence 4, Appli

C 13	62.5	37.0	2155	4	US-09-639-696C-4	Sequence 4, Appli
C 14	62.5	37.0	4480	3	US-09-191-171-7	Sequence 7, Appli
C 15	62.5	37.0	4480	3	US-09-385-707-7	Sequence 7, Appli
C 16	62.5	37.0	6200	4	US-09-439-923-1	Sequence 1, Appli
C 17	62.5	37.0	6200	4	US-09-711-202A-1	Sequence 1, Appli
C 18	62.5	37.0	6200	4	US-09-711-205A-1	Sequence 1, Appli
C 19	62.5	37.0	6238	4	US-09-639-696C-6	Sequence 6, Appli
C 20	62.5	37.0	169998	4	US-09-676-610B-24	Sequence 24, Appli
C 21	62.5	37.0	17496	4	US-09-877-177A-10	Sequence 10, Appli
C 22	62	36.7	365	4	US-09-621-976-11459	Sequence 11459, A
C 23	62	36.7	825	4	US-09-489-039A-5870	Sequence 5870, Ap
C 24	62	36.7	2502	4	US-09-489-039A-1765	Sequence 1765, Ap
C 25	62	36.7	3780	4	US-09-489-039A-1669	Sequence 1669, Ap
C 26	61.5	36.4	702	4	US-09-252-991A-1115	Sequence 1115, Ap
C 27	61.5	36.4	2178	4	US-09-252-991A-1039	Sequence 1039, Ap
C 28	61.5	36.4	2433	4	US-09-252-991A-923	Sequence 923, App
C 29	61	36.1	972	4	US-09-489-039A-2993	Sequence 2993, Ap
C 30	61	36.1	2502	4	US-09-069-023-2	Sequence 2, Appli
C 31	60.5	35.8	657	4	US-09-489-039A-2986	Sequence 2986, Ap
C 32	60.5	35.8	1452	4	US-09-489-039A-2735	Sequence 2735, Ap
C 33	60.5	35.8	43280	2	US-08-804-227C-1	Sequence 1, Appli
C 34	60	35.5	894	4	US-09-489-039A-5631	Sequence 5631, Ap
C 35	60	35.5	1113	4	US-09-489-039A-5264	Sequence 5264, Ap
C 36	60	35.5	1208	2	US-08-403-852D-4	Sequence 4, Appli
C 37	60	35.5	1208	3	US-08-510-646B-4	Sequence 4, Appli
C 38	60	35.5	1208	3	US-09-231-818-4	Sequence 4, Appli
C 39	60	35.5	1208	4	US-09-635-359B-4	Sequence 4, Appli
C 40	60	35.5	1611	4	US-09-808-701A-13	Sequence 13, Appli
C 41	60	35.5	2126	2	US-08-789-354-1	Sequence 1, Appli
C 42	60	35.5	2126	3	US-09-110-937-1	Sequence 1, Appli
C 43	60	35.5	2126	3	US-09-058-725B-1	Sequence 1, Appli
C 44	60	35.5	2126	3	US-09-232-857-1	Sequence 1, Appli
C 45	60	35.5	5392	2	US-08-403-852D-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-252-991A-9484  
Sequence 9484, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubinfeld et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 9484  
LENGTH: 438  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9484

## Alignment Scores:

Pred. No.: 33.2  
Score: 64.00  
Percent Similarity: 45.24%  
Best Local Similarity: 40.48%  
Query Match: 37.87%  
DB: 4  
Gaps: 3

US-10-010-050A-2\_COPY\_1\_30 (1-30) x US-09-252-991A-9484 (1-438)

Qy 2 ArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTP-----14  
DB 5 CGCGACCCCGCTGGCGCGCTCGCAGCGCGGCGGATGCGCGTATCGACAGCAGCAGG 64









NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7956-0011-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2067 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
IS-08-713-928B-8

Alignment Scores:  
Pred. No.: 268 Length: 2067  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 36.98% Indels: 11  
DB: 2 Gaps: 1

US-10-010-050A-2\_COPY\_1\_30 (1-30) x US-08-713-928B-8 (1-2067)

Y 2 ArgArgG1yAla-----G1yAlaAlaArgG1y 10  
b 1081 CGAAGGGGTGCGGTGTAGCTCAGAGGCAATTGCTGCTCAGAGGCGCTAGGGGA 1022  
Y 11 ArgAlaSerTTPCySTTPAlaLeuAla 19  
b 1021 AGGCGGAGGTGGTGTGTTGCCACAGTAGCA 995

RESULT 11  
S-09-191-171-4/c  
Sequence 4, Application US/09191171  
Patent No. 6149909  
Patent No. 6149909 6143294  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
TITLE OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/191,171  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/494,104  
FILING DATE: 23-JUN-1995  
APPLICATION NUMBER: AU PK9490/91  
FILING DATE: 14-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PCT/AU92/00611

FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/084,254  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: DIGILIO, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 89782  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2155 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 89..2047  
US-09-191-171-4

Alignment Scores:  
Pred. No.: 281 Length: 2155  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 36.98% Indels: 11  
DB: 3 Gaps: 1

US-10-010-050A-2\_COPY\_1\_30 (1-30) x US-09-191-171-4 (1-2155)

QY 2 ArgArgG1yAla-----G1yAlaAlaArgG1y 10  
Db 1169 CGAAGGGGTGCGGTGTAGCTCAGAGGCAATTGCTGCTCAGAGGCGCTAGGGGA 1110  
QY 11 ArgAlaSerTTPCySTTPAlaLeuAla 19  
Db 1109 AGGCGGAGGTGGTGTGTTGCCACAGTAGCA 1083

RESULT 12  
US-09-385-707-4/c  
Sequence 4, Application US/09385707  
Patent No. 6238662  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
TITLE OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,707  
FILING DATE:  
CLASSIFICATION:

1. *Journal of the American Medical Association*, 1997; 277: 1001-1005.

1. **Identify the main components of the system.** The system consists of a **client** and a **server**. The client is responsible for sending requests to the server, and the server is responsible for processing these requests and returning responses.

10  
1110

$\frac{1}{2} \left( \frac{1}{2} + \frac{1}{2} \right) = \frac{1}{2}$

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/191,171  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/494,104  
FILING DATE: 23-JUN-1995  
APPLICATION NUMBER: AU PK9490/91  
FILING DATE: 14-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PCT/AU92/00611  
FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/084,254  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4480 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-191-171-7

Alignment Scores:  
Pred. No.: 627 Length: 4480  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 36.98% Indels: 11  
Gaps: 1

US-10-010-050a-2\_COPY\_1\_30 (1-30) x US-09-191-171-7 (1-4480)

Y 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10  
|||  
b 2192 CGAAGGGGTGGCGGTGAGCTCAGGAAGGCAATTGCTGCTCAGAGCGGCTAGGGGA 2133  
|||  
Y 11 ArgAlaSerTrpCysTrpAlaLeuAla 19  
|||  
b 2132 AGGCGGAGGTGGTGTGGCCAGTAGCA 2106  
|||

RESULT 15  
US-09-385-707-7/c  
Sequence 7, Application US/09385707  
Patent No. 6238662  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
TITLE OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York

COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,707  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/494,104  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4480 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-385-707-7

Alignment Scores:  
Pred. No.: 627 Length: 4480  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 36.98% Indels: 11  
Gaps: 1

US-10-010-050a-2\_COPY\_1\_30 (1-30) x US-09-385-707-7 (1-4480)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10  
|||  
Db 2192 CGAAGGGGTGGCGGTGAGCTCAGGAAGGCAATTGCTGCTCAGAGCGGCTAGGGGA 2133  
|||  
QY 11 ArgAlaSerTrpCysTrpAlaLeuAla 19  
|||  
Db 2132 AGGCGGAGGTGGTGTGGCCAGTAGCA 2106  
|||

Search completed: April 25, 2004, 05:51:41  
Job time: 12.9583 secs







proteins - used for diagnosis and treatment of thyroid disorders or diseases involving genes on chromosome 13.

Claim 27, Page 88-90; 101pp; English.

This invention describes the isolation of a novel human secreted protein, zsi946 encoded by a gene on chromosome 13 which is mainly expressed in the thyroid. This product can be used to study secretion of proteins from cells and also to treat or prevent deficient expression of zsi946, which may be associated with thyroid diseases (e.g. hypothyroidism, Graves' disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that involve genes in the same region of chromosome 13 (e.g. Hirschsprung's disease, neuronal ceroid-lipofuscinosis, Wilson disease and Reiter syndrome). Antibodies and other binding proteins, are used as immunoassay reagents to detect zsi946 or cells expressing it, e.g. for assessing thyroid function to produce anti-idiotypic antibodies, for affinity purification of zsi946, to screen expression libraries, to neutralise zsi946 activity, and to deliver toxins, radioisotopes etc. for therapeutic or diagnostic purposes. Agonists of the product can be used to promote growth, differentiation and proliferation of specific cell types, e.g. for treating (extra)thyroid diseases or as additive to cell cultures

Sequence 1486 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;

Alignment Scores:

Prod. No.:	Length:	1486
169.00	Matches:	30
100.00%	Conservative:	0
100.00%	Mismatches:	0
100.00%	Indels:	0
2	Gaps:	0

US-10-010-050A-2\_COPY\_1\_30 (1-30) x AA02855 (1-1486)

1 MetArgArgGlyAlaGlyAlaAlaArgGlyAlaSerTrpCysTrpAlaLeuAlaLeu 20

47 ATGGCGGCGGGCGGGCGGGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 106

21 LeuTrpLeuAlaValAlaProGlyTrpSer 30

107 CTTGGCTCGCGGTGCTCCGGGCTGCTC 136

RESULT 4

AAZ24826

AAZ24826 standard; DNA; 1751 BP.

AAZ24826;

02-DEC-1999 (first entry)

Human secreted protein gene 16 clone HMZAD77.

Human; secreted protein; fusion protein; gene therapy; protein therapy; diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; developmental abnormality; foetal deficiency; blood; allergy; renal; ds; immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS; cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion; endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

MO9947540-A1.

23-SEP-1999.

18-MAR-1999; 99WC-US005804.

19-MAR-1998; 98US-0078563P.

19-MAR-1998; 98US-0078566P.

19-MAR-1998; 98US-0078573P.

19-MAR-1998; 98US-0078574P.

19-MAR-1998; 98US-0078576P.

19-MAR-1998; 98US-0078577P.

19-MAR-1998; 98US-0078578P.

19-MAR-1998; 98US-0078579P.

19-MAR-1998; 98US-0078581P.

19-MAR-1998; 98US-0080312P.

01-APR-1998; 98US-0080313P.

01-APR-1998; 98US-0080314P.

(HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Nt J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR; Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Latleur DW, Olsen HS; Shi Y, Moore PA; WPI; 1999-562050/47. P-PSDB; AAY41323.

New isolated human genes, useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders.

Claim 1; Page 308; 484pp; English.

This sequence represents a nucleic acid molecule which encodes a secreted human protein. The gene number, and the clone it is derived from, are detailed in the descriptor line. The gene can be used to generate fusion proteins by linking to the gene to a human immunoglobulin Fc portion (e.g. AAZ24802) for increasing the stability of the fused protein as compared to the human protein only. The invention relates to 95 novel CC genes and their fragments (nucleic acid sequences: AAZ24811-Z24907; amino acid sequences AAAY1308-V41404) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, CC pathological conditions can be diagnosed by determining the amount of the CC new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the CC 95 polynucleotides, based on which tissues they are most highly expressed in (see AAZ24811 for described uses)

Sequence 1751 BP; 477 A; 355 C; 380 G; 523 T; 0 U; 16 Other;

Alignment Scores:

Prod. No.:	Length:	1751
169.00	Matches:	30
100.00%	Conservative:	0
100.00%	Mismatches:	0
100.00%	Indels:	0
2	Gaps:	0

US-10-010-050A-2\_COPY\_1\_30 (1-30) x AAZ24826 (1-1751)

1 MetArgArgGlyAlaGlyAlaAlaArgGlyAlaSerTrpCysTrpAlaLeuAlaLeu 20

49 ATGCGCGGGCGGGCGGGCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108

21 LeuTrpLeuAlaValAlaProGlyTrpSer 30

109 CTTGGCTCGCGGTGCTCCGGGCTGCTC 138

RESULT 5

AAK94829

AAK94829 standard; cDNA; 2120 BP.

AAK94829;

06-NOV-2001 (first entry)

Human full-length cDNA, SEQ ID NO: 3977.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

PX	FN	EP1130094-A2.
KX	PD	05-SEP-2001.
KX	PF	07-JUL-2000; 2000EP-00114089.
KX	PR	08-JUL-1999; 99JP-00194486.
KX	PR	11-JAN-2000; 2000JP-00118774.
KX	PR	02-MAY-2000; 2000JP-00183765.
PA	(HELI-) HELIX RES INST.	
KX	Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
PI	Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
KX	WPI; 2001-524255/58.	
DR	P-PDSB; AAM93870.	
PT	830 Primers useful for synthesizing full length cDNA clones and their use	
XX	in genetic manipulation.	
PS	Claim 8; SEQ ID NO 397; 1380bp + Sequence Listing; English.	
KX	The invention relates to primers for synthesising full length cDNA	
CC	clones. 830 cDNA molecules encoding a human protein have been isolated	
CC	and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have	
CC	been determined. Primers for synthesising the full length cDNA are useful	
CC	for clarifying the function of the protein encoded by the cDNA. The full	
CC	length clones were obtained by construction of full length enriched cDNA	
CC	libraries that were synthesised by the oligo-capping method. The primers	
CC	enable the production of the full length cDNA easily without any special	
CC	methods. The present sequence is a full length human cDNA of the	
CC	invention. Note: The sequence data for this patent did not form part of	
CC	the printed specification, but was obtained in CD-ROM format directly	
CC	from EPO	
XX	Sequence 2120 BP; 590 A; 409 C; 460 G; 661 T; 0 U; 0 Other;	
SQ		
Alignment Scores:		
Pred. No.:	1.23e-09	Length: 2120
Score:	169.00	Matches: 30
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	4	Gaps: 0
US-10-010-050A-2_COPY_1_30 (1-30) x AA694829 (1-2120)		
DY	1 MetArgArgGlyAlaIValaIValaIaaArgGlyArgGlaIsertProCysTPrAlaLeuAlaIeu 20	
Db	52 ATGGGGGGGGGCCCGGGCCCGGGCTCGGGAGACCGGCTCTGTGTCGTGGGCCCCCTGCCTG 111	
DY	21 LeuTrpleuAlaValaIalProGlyITPser 30	
Db	112 CTTTGCGTGGCGGTGGTTCGGGGCTGTGTC 141	
RESULT 6		
ABQ40653/c		
ID	ABQ40653 standard; DNA; 630 BP.	
XX	ABQ40653;	
XX	12--JUL-2002 (first entry)	
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 27244.	
XX	Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;	
KW	drug; side effect; cancer; central nervous system; cardiovascular;	
KM	gastrointestinal; respiratory system; single nucleotide polymorphism;	
KM	SNP; cell differentiation; de.	
XX	Homo sapiens.	
OS		

[illegible]









```

R      11-JAN-2000; 2000US-00480902.
X      (BYAT/) BYATT J C.
X      (MATH/) MATHIALAGAN N.
X      (TAON/) TAO N.
X      (WARR/) WARREN W C.
X      Byatt JC, Mathialagan N, Tao N, Warren WC;
X      WPI; 2003-110599/10.
X      New nucleic acid associated with lactation, and muscle and fat
X      deposition, useful for genome mapping, gene identification and analysis,
X      cattle breeding, or for genetically improving cattle.
X      Claim 2; SEQ ID NO 6089; 245bp; English.
X      The invention relates to a purified nucleic acid molecule associated with
X      lactation or muscle and fat deposition (designated LMFD), derived from
X      cattle, and the LMFD nucleic acid can specifically hybridise to a second
X      nucleic acid molecule comprising any of 15112 nucleotide sequences,
X      appearing as ABX34836-ABX49947, or complements of them. Also included are
X      (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
X      acid linked to a promoter and a 3' non-translated sequence that
X      functions in the cell to cause termination of transcription and addition
X      of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
X      (2) determining a level or pattern of a molecule in a bovine cell or
X      tissue comprising: (a) incubating a marker nucleic acid (comprising any
X      of the 15112 nucleic acid sequences or its complement or fragment) with a
X      complementary nucleic acid molecule obtained from the bovine cell or
X      tissue, where hybridisation between the marker nucleic acid and the
X      complementary nucleic acid permits the detection of the molecule; and (b)
X      detecting the level or pattern of the complementary nucleic acid, where
X      the detection of the complementary nucleic acid is predictive of the
X      level or pattern of the molecule. The LMFD nucleic acid is used for
X      determining a level or pattern of a molecule in a bovine cell or tissue.
X      It is useful for genome mapping, gene identification and analysis, cattle
X      breeding, preparation of constructs for use in cattle gene expression, or
X      for genetically improving cattle. The present sequence is one of the
X      15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
X      present sequence was not shown in the specification but was obtained in
X      electronic format from the USPTO web site:
X      seqdata.uspto.gov/Sequence.html?DocID=20020137139
X      Sequence 403 BP; 90 A; 130 C; 114 G; 69 T; 0 U; 0 Other;
X      Alignment Scores:
X      Aligned. No.: 31.3 Length: 403
X      Score: 69.50 Matches: 15
X      Percent Similarity: 48.48% Conservative: 1
X      Best Local Similarity: 45.45% Mismatches: 10
X      Query Match: 41.12% Indels: 7
X      B: 7 Gaps: 1
X      JS-10-010-050A-2 COPY_1_30 (1-30) X ABX40924 (1-403)
X      Y      4 GYALAGIYALALALAGIYARGIYALaSerTTCys-----TTP 16
X      Y      283 GAGGACAGGTGCATCAGTGTCTCTGTGCATCTTCGACCTGGTGCGCTTTCATGG 224
X      Y      17 AAlaenAlaIeuIeuTTPLeuAlaValValProGlyTTP 29
X      Y      223 GCTGTGGGCGCTGCATGGGACAGCGCCGCGATCAGCATGG 185
X      RESULT 15
X      ABX46110/c
X      ID ABX46110 standard; cDNA; 404 BP.
X      AC ABX46110;
X      XX
X      XT 21-FEB-2003 (first entry)
X      ZE Bovine EST associated with lactation/muscle/fat deposition #11275.

```

XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
KM muscle deposition; fat deposition; genome mapping; gene identification;  
XX gene analysis; cattle breeding.  
XX  
XX Bos Taurus.  
XX  
XX US2002137139-A1.  
PN  
XX  
XX 26-SEP-2002.  
PD  
XX  
XX 24-SEP-2001; 2001US-00360352.  
PF  
XX  
XX 12-JAN-1999; 99US-0115707P.  
PR 11-JAN-2000; 2000US-00480902.  
RR  
XX  
PA (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
PI Byatt JC, Mathialagan N, Tao N, Warren WC,  
DR WPI, 2003-110599/10.  
XX  
PT New nucleic acid associated with lactation, and muscle and fat  
PT deposition, useful for genome mapping, gene identification and analysis,  
PT cattle breeding, or for genetically improving cattle.  
PS  
XX Claim 2; SEQ ID NO 11275; 245bp; English.

The invention relates to a purified nucleic acid molecule associated with lactation or muscle and fat deposition (designated LMFD), derived from cattle, and the LMFD nucleic acid can specifically hybridise to a second nucleic acid molecule comprising any of 15112 nucleotide sequences, appearing as ABX34836-ABX49947, or complements of them. Also included are:  
(1) a transformed cell having a nucleic acid comprising an LMFD nucleic acid linked to a promoter and a 3' non-translated sequence that functions in the cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and  
(2) determining a level or pattern of a molecule in a bovine cell or tissue comprising: (a) incubating a marker nucleic acid (comprising any of the 15112 nucleic acid sequences or its complement or fragment) with a complementary nucleic acid molecule obtained from the bovine cell or tissue, where hybridisation between the marker nucleic acid and the complementary nucleic acid permits the detection of the molecule; and (b) detecting the level or pattern of the complementary nucleic acid, where the detection of the complementary nucleic acid is predictive of the level or pattern of the molecule. The LMFD nucleic acid is used for determining a level or pattern of a molecule in a bovine cell or tissue. It is useful for genome mapping, gene identification and analysis, cattle breeding, preparation of constructs for use in cattle gene expression, or for genetically improving cattle. The present sequence is one of the 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The present sequence was not shown in the specification but was obtained in electronic format from the USPTO web site:  
[seqdata.uspto.gov/sequence.html?docid=20020137139](http://seqdata.uspto.gov/sequence.html?docid=20020137139)

Sequence 404 BP; 91 A; 131 C; 115 G; 67 T; 0 U; 0 Other;

Alignment Scores:

Pred. No. :	Length:	Score:	Matches:	Conservative:	Mismatches:	Best Local Similarity:	Indels:
31.4	404	69.50	15	1	10	45.45*	7
41.12*							1

Query Match: 7 Gaps: 1

DB: DB

US-10-010-050A-2\_COPY\_1\_30 (1-30) x ABX46110 (1-404)

QY 4 G|YA|AG|YA|AA|A|AGG|YA|A|GA|A|AG|T|P|C|Y|S|-----TTP 16  
|||:::|||||  
304 GGAACGAGCGTCATCAACGTGGCTCCGTGCATCATGTTGCCGACTTGAGTGGCGCTTCATGG 245





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:02:32 ; Search time 355.406 Seconds  
(without alignments)  
2352.639 Million cell updates/sec

Title: US-10-010-050a-2\_COPY\_1\_28  
Percent score: 154  
Sequence: 1 MRRGAGARGRASWCALALLMLAVPG 28

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-O/cp2 1/USPTO.spool/US10010050/runat.22042004.113204.27556/app.query.fasta\_1.1372  
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human0.cdt -ITR=45  
-LOCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US10010050 @CEN 1 1 6283 @runat.22042004.113204.27556 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUTERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pyo:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_plg:\*  
27: em\_gss\_vit:\*  
28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	* Match	Length	DB	ID	Description
1	154	100.0	565	9	AU280382	AU280382 AU280382
2	154	100.0	627	12	BM849262	BM849262 K-EST0129
3	154	100.0	623	13	BQ417448	BQ417448 IK3805.Y
4	154	100.0	676	12	BG820113	BG820113 602782325
5	154	100.0	787	12	B1223533	B1223533 602941923
6	154	100.0	847	13	BQ719074	BQ719074 AGENCOURT
7	154	100.0	869	12	BI090566	BI090566 602855673
8	154	100.0	923	13	BX331615	BX331615 BX331615
9	154	100.0	1098	10	BF982158	BF982158 602308976
10	154	100.0	1201	9	AL546472	AL546472 AL546472
11	154	100.0	1222	13	BQ897986	BQ897986 AGENCOURT
12	146	94.8	924	14	CA488543	CA488543 AGENCOURT
13	141	91.6	1077	13	BX342662	BX342662 BX342662
14	127	82.5	589	10	BF977311	BF977311 602146485
15	124	80.5	555	14	CA397574	CA397574 csn2h10.Y
16	104	67.5	273	9	AA300650	AA300650 EST13763
17	73	47.4	383	14	CB808908	CB808908 AMGNNUC.M
18	73	47.4	424	14	CB795968	CB795968 AMGNNUC.M
19	73	47.4	427	14	CB548129	CB548129 AMGNNUC.M
20	72	46.8	1355	10	BE880806	BE880806 601493233
21	71	46.1	569	29	CE827309	CE827309 tigr-gss-
22	71	46.1	532	13	BX844586	BX844586 BX844586
23	69.5	45.1	1258	13	BQ945568	BQ945568 AGENCOURT
24	69	44.8	367	13	BY014529	BY014529 BY014529
25	69	44.8	377	13	BY015054	BY015054 BY015054
26	69	44.8	395	13	BY027225	BY027225 BY027225
27	69	44.8	421	13	BY247307	BY247307 BY247307
28	69	44.8	451	10	BB639395	BB639395 BB639395
29	69	44.8	665	13	B7721389	B7721389 BY721389
30	69	44.8	686	14	CB876342	CB876342 ttr1c041xb
31	69	44.8	703	14	CB676342	CB676342 ttr1c041xb
32	69	44.8	722	14	CB660119	CB660119 OSJNED01T
33	69	44.8	771	14	CB905022	CB905022 ttr1c041xb
34	69	44.8	796	12	B1182240	B1182240 UNL-P-FN-
35	68.5	44.5	571	12	BM171756	BM171756 imageqc.5
36	68.5	44.5	642	10	BF900979	BF900979 IL0-MT030
37	68.5	44.5	977	10	BF032072	BF032072 601559649
38	68	44.2	743	29	CNS01LHE	AL149635 Anopheles
39	68	44.2	773	28	BZ733787	BZ733787 OGDSD56TM
40	68	44.2	827	29	CNS01RBN	AL157095 Anopheles
41	67.5	43.8	411	13	B1227424	BY227424 BY227424
42	67.5	43.8	561	9	A1786906	A1786906 UJ32909.Y
43	67.5	43.8	587	10	BG077284	BG077284 H3013B06-
44	67.5	43.8	789	29	CG276043	CG276043 OGWGFS2TV
45	67.5	43.8	825	12	BG536228	BG536228 602565449

#### ALIGNMENTS

RESULT 1  
LOCUS AU280382 565 bp mRNA linear EST 31-JUL-2003  
DEFINITION AU280382 NIBSE2 Homo sapiens cDNA clone NIBSE2001763 5', mRNA  
SEQUENCE.  
ACCESSION AU280382  
VERSION AU280382.1 GI:28299609  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 565)





ORIGIN  
 /clone\_1ib="NIH\_MGC\_12"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."

## ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3.55e-06	787	154.00	28	0	0	0
Score:	100.00%	100.00%	0	0	0	0
Percent Similarity:	100.00%	100.00%	0	0	0	0
Best Local Similarity:	100.00%	100.00%	0	0	0	0
Query Match:	100.00%	100.00%	0	0	0	0
DB:	12	Gaps:	0	0	0	0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x B1223533 (1-787)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20  
 |||  
 DB 26 ATGCGGCGGCGGCGGCGGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTG 85

QY 21 LeuTrpLeuAlaValAlaProGly 28  
 |||  
 DB 86 CTTGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTG 109

RESULT 6 847 bp mRNA linear EST 16-JUL-2002  
 BQ719074  
 LOCUS AGENCOURT 8103813 lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
 DEFINITION IMAGE:6191057 5', mRNA sequence.  
 ACCESSION BQ719074  
 VERSION BQ719074.1 GI:21857971  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 847)  
 NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM13590 row: n column: 18  
 High quality sequence stop: 484.  
 Location/Qualifiers

FEATURES  
 source  
 1..847  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6191057"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /clone\_1ib="lupski\_sympathetic\_trunk"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:  
 NotI; Site 2: SalI; cDNA made by oligo-dT priming.  
 directionally cloned using the following adaptors:  
 5'-TCGACCAAGCGCTCCG-3' and  
 5'-GACTAGTTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >  
 1 kb for average insert length 1.9 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (aylor  
 College of Medicine); available through Life  
 Technologies."

## ORIGIN

## ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3.9e-06	847	154.00	28	0	0	0
Score:	100.00%	100.00%	0	0	0	0
Percent Similarity:	100.00%	100.00%	0	0	0	0
Best Local Similarity:	100.00%	100.00%	0	0	0	0
Query Match:	100.00%	100.00%	0	0	0	0
DB:	13	Gaps:	0	0	0	0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x BQ719074 (1-847)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20  
 |||  
 DB 80 ATGCGGCGGCGGCGGCGGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTG 139

QY 21 LeuTrpLeuAlaValAlaProGly 28  
 |||  
 DB 140 CTTGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTG 163

RESULT 7 869 bp mRNA linear EST 20-JUN-2001  
 B1090566  
 LOCUS 602853673F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:4996891 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1090566  
 VERSION B1090566.1 GI:14508896  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 869)  
 NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM11023 row: a column: 20  
 High quality sequence stop: 843.  
 Location/Qualifiers

FEATURES  
 source  
 1..869  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4996891"  
 /cell\_line="MGC36"  
 /lab\_host="DH10B"  
 /lab\_1ib="NIH\_MGC\_10"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.5 kb. Library prepared by Life  
 Technologies."

## ORIGIN

## ALIGNMENT SCORES:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.03e-06	869	154.00	28	0	0	0
Score:	100.00%	100.00%	0	0	0	0
Percent Similarity:	100.00%	100.00%	0	0	0	0
Best Local Similarity:	100.00%	100.00%	0	0	0	0
Query Match:	100.00%	100.00%	0	0	0	0
DB:	12	Gaps:	0	0	0	0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x B1090566 (1-869)

QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20









Best Local Similarity: 96.43% Mismatches: 1  
 Query Match: 82.47% Indels: 1  
 DB: 10 Gaps: 0

US-10-010-050a-2\_copy\_1\_28 (1-28) x BP977311 (1-589)

Qy 1 MetArGArGlyAlaGlyAlaAlaArgGlyArGAlaSerTrpCySTrPAlaLeuAlaLeu 20  
 Db 161 ATGGCGCGGGCGCGCGCGCGGT-CGGGAGCGCGCTTCTGTCGTGCGCCCTGCGCTG 2.9

Qy 21 LeuTrpLeuAlaAlaAlaProGly 28  
 Db 220 CTTTGCTCGCGGTGCTCCGGC 243

RESULT 15  
 CA397574 555 bp mRNA linear EST 06-NOV-2002  
 LOCUS cs92h10.y2 Human Retinal pigment epithelium/choroid cDNA  
 DEFINITION (Un-normalized, unamplified): cs Homo sapiens cDNA clone cs92h10  
 5', mRNA sequence.

ACCESSION CA397574  
 VERSION CA397574.1 GI:24735005  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 555)  
 Wistow,G., Bernstein,S.L., Wyatt,M.K., Farris,R.N., Behal,A.,  
 Touchman,J.W., Bouffard,G., Smith,D. and Peterson,K.  
 Expressed sequence tag analysis of human RPE/choroid for the  
 NEIBank Project: Over 6000 non-redundant transcripts, novel genes  
 and splice variants  
 Mol. Vis. 8 (4), 205-220 (2002)

JOURNAL MEDLINE  
 PUBMED  
 COMMENT 12107410  
 22103460

Section on Molecular Structure and Function  
 National Eye Institute  
 6/331, NIH, Bethesda, MD 20892-2740, USA  
 Tel: 301 402 3452  
 Fax: 301 496 0078  
 Email: graeme@helix.nih.gov

Plate: 92 row: h column: 10  
 Seq primer: M13RP1 reverse primer (ABI).  
 Location/Qualifiers

FEATURES  
 source 1..555

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="cs92h10"  
 /tissue\_type="RPE/choroid"  
 /dev\_stage="Adult"  
 /lab\_host="EMDH10B"  
 /clone\_lib="Human Retinal pigment epithelium/choroid cDNA  
 (Un-normalized, unamplified): cs"  
 /note="Organ: Eye; Vector: PCMVSPORT6; Two different donor  
 eyes (75-80 years old) yielded approximately 600 mg of  
 dissected RPE/choroid tissue. This in turn yielded 340 ug  
 of total RNA and 7 ug of mRNA. A directionally cloned cDNA  
 library in the PCMVSPORT6 vector was constructed at Life  
 Technologies (Rockville, MD; now part of Invitrogen Corp),  
 essentially following the protocols of the Superscript  
 Plasmid System (Invitrogen Corp  
 <http://www.invitrogen.com/>). The library code  
 designation was cs. For this library, cDNA inserts were  
 cloned into the NotI/Mlu sites of the vector. EST  
 analysis was performed on the unamplified library at the  
 NIH Intramural Sequencing Center (NISC)."

ORIGIN  
 Alignment Scores:  
 Pred. No.:

0.00288

Length:

555

Score: 124.00 Matches: 23  
 Percent Similarity: 82.14% Conservative: 0  
 Best Local Similarity: 82.14% Mismatches: 5  
 Query Match: 80.52% Indels: 0  
 DB: 14 Gaps: 0

US-10-010-050a-2\_copy\_1\_28 (1-28) x CA397574 (1-555)

Qy 1 MetArGArGlyAlaGlyAlaAlaArgGlyArGAlaSerTrpCySTrPAlaLeuAlaLeu 20  
 Db 40 ATGGCGCGGGCGCGCGCGCGGT-CGGGAGCGCGCTTCTGTCGTGCGCCCTGCGCTG 99

Qy 21 LeuTrpLeuAlaAlaProGly 28  
 Db 100 CTTTGCTCGCGGTGCTCCGGC 123

Search completed: April 25, 2004, 05:46:19  
 Job time: 361.406 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

M protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:04:07 ; Search time 55.8056 Seconds

(without alignments)  
2262.168 Million cell updates/sec

Title: US-10-010-050a-2\_COPY\_1\_28  
Perfect score: 154  
Sequence: 1 MRGAGAGRGRASWCALALLMLAVVPG 28

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2907579 seqs, 2254313464 residues

Total number of hits satisfying chosen parameters: 5815158

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US10010050/rnphat.22042004.113205.27584/app.query.fasta\_1.1372  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rmph -MINMATCH=0.1  
-LOPCU=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cgi -LIST=45 -DOCCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US10010050 @cgn\_1\_1\_776 @rnphat.22042004.113205.27584  
-NCPU=6 -ICPU=3 -NO MAP -LARGOUDERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications NA:\*

1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/2/pubpna/PCF\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/2/pubpna/PCFUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cgn2\_6/prodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
15: /cgn2\_6/prodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
16: /cgn2\_6/prodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
19: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	154	100.0	1486	9	US-09-122-383-1	Sequence 1, Appl
2	154	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appl
3	154	100.0	1751	13	US-10-653-595-26	Sequence 26, Appl
4	154	100.0	1751	13	US-09-397-945-26	Sequence 26, Appl
5	106	68.8	1038	14	US-09-122-383-13	Sequence 13, Appl
6	106	68.8	1038	14	US-10-010-050A-13	Sequence 13, Appl
7	66	42.9	7155	15	US-10-329-079-14	Sequence 14, Appl
8	66	42.9	37350	15	US-10-329-079-14	Sequence 14, Appl
9	64	41.6	580	13	US-10-027-632-61	Sequence 24, Appl
10	64	41.6	580	16	US-10-027-632-241716	Sequence 24, Appl
11	64	41.6	1071	13	US-10-282-122A-15190	Sequence 24, Appl
12	63.5	41.2	276276	13	US-10-087-192-754	Sequence 754, App
13	63	40.9	772	15	US-10-029-517-106	Sequence 106, App
14	63	40.9	1257	13	US-10-425-114-5227	Sequence 5227, App
15	63	40.9	322101	13	US-10-354-247-1	Sequence 1, Appl
16	63	40.9	322101	15	US-10-060-902-1	Sequence 1, Appl
17	62.5	40.6	256	9	US-09-960-352-12900	Sequence 12900, A
18	62.5	40.6	1872	10	US-09-764-891-7060	Sequence 7060, App
19	62.5	40.6	1962	13	US-10-411-037-65	Sequence 65, Appl
20	62.5	40.6	1962	13	US-10-411-037-65	Sequence 65, Appl
21	62.5	40.6	2155	9	US-09-962-436-299	Sequence 299, App
22	62.5	40.6	2155	9	US-09-880-107-2394	Sequence 2394, App
23	62.5	40.6	6200	9	US-09-993-038-1	Sequence 1, Appl
24	62.5	40.6	6200	9	US-09-993-038-1	Sequence 1, Appl
25	62.5	40.6	6200	15	US-10-206-443-1	Sequence 1, Appl
26	62.5	40.6	7648	17	US-10-176-066-1	Sequence 1, Appl
27	62.5	40.6	169998	16	US-10-380-931-24	Sequence 24, Appl
28	62.5	40.6	197496	9	US-09-877-177-10	Sequence 10, Appl
29	62.5	40.6	197496	13	US-10-426-836-10	Sequence 10, Appl
30	62	40.3	566	10	US-09-930-213-116	Sequence 116, App
31	62	40.3	2024	16	US-10-264-237-1269	Sequence 1269, App
32	61.5	39.9	404	9	US-09-960-352-6089	Sequence 6089, App
33	61.5	39.9	404	9	US-09-960-352-11275	Sequence 11275, A
34	61.5	39.9	414	9	US-09-960-352-14095	Sequence 14095, A
35	61.5	39.9	988	13	US-10-425-114-30764	Sequence 30764, A
36	61	39.6	558	13	US-10-027-632-67820	Sequence 67820, A
37	61	39.6	558	13	US-10-027-632-67821	Sequence 67821, A
38	61	39.6	558	13	US-10-027-632-67822	Sequence 67822, A
39	61	39.6	558	16	US-10-027-632-67820	Sequence 67820, A
40	61	39.6	558	16	US-10-027-632-67821	Sequence 67821, A
41	61	39.6	558	16	US-10-027-632-67822	Sequence 67822, A
42	61	39.6	96595	12	US-09-997-722-262	Sequence 262, App
43	61	39.6	684973	9	US-09-263-959-1	Sequence 1, Appl
44	60.5	39.3	596	9	US-09-811-284-39	Sequence 39, Appl
45	60.5	39.3	2506	13	US-10-282-122A-20063	Sequence 20063, A

## ALIGNMENTS

RESULT 1  
US-09-122-383-1  
Sequence 1, Application US/09122383A  
Patent No. US20020042093A1  
GENERAL INFORMATION:  
APPLICANT: Shepard, Paul O.  
APPLICANT: Gilbertson, Debra G.  
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
FILE REFERENCE: CHROMOSOME 13  
CURRENT FILING DATE: 1997-07-24  
EARLIER APPLICATION NUMBER: US/09/122,383A  
EARLIER APPLICATION NUMBER: 60/053,613  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 1486  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (47)...(1084)

US-09-122-383-1

## Alignment Scores:

Pred. No.:	8.32e-11	Length:	1486
Score:	154.00	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-09-122-383-1 (1-1486)

QY 1 MetargarglYAlaGlyAlaAlaarglYArGAlaSerTPCySTrPAlaLeuAlaLeu 20

Db 47 ATGCCGCGGGCGCGCGCGCGCTCGGGACGCGCTTCTGTGCTGGGCGCGCTG 136

QY 21 LeutrpLeuAlaValaValaProgly 28

Db 107 CTTGGCTCGCGGTGGTTCGGGCG 130

## RESULT 2

US-10-010-050A-1

Sequence 1, Application US/10010050A  
Publication No. US20020173624A1

GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.

APPLICANT: Gilbertson, Debra G.

TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN

FILE REFERENCE: 97-38C1

CURRENT APPLICATION NUMBER: US/10/010,050A

PRIOR FILING DATE: 2002-03-26

PRIOR APPLICATION NUMBER: US 09/122,383

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 60/053,613

PRIOR FILING DATE: 1997-07-24

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1486

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: CDS

LOCATION: (47)...(1084)

US-10-010-050A-1

## Alignment Scores:

Pred. No.:	8.32e-11	Length:	1486
Score:	154.00	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-10-010-050A-1 (1-1486)

QY 1 MetargarglYAlaGlyAlaAlaarglYArGAlaSerTPCySTrPAlaLeuAlaLeu 20

Db 47 ATGCCGCGGGCGCGCGCGCGCTCGGGACGCGCTTCTGTGCTGGGCGCGCTG 106

QY 21 LeutrpLeuAlaValaValaProgly 28

Db 107 CTTGGCTCGCGGTGGTTCGGGCG 130

## RESULT 3

US-10-653-595-26

Sequence 26, Application US/10653595  
Publication No. US20040048304A1

GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

FILE REFERENCE: P2027P1C1

CURRENT APPLICATION NUMBER: US/10/653,595

CURRENT FILING DATE: 2003-09-03

PRIOR APPLICATION NUMBER: US 09/397945

PRIOR FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: 60/078,566

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,576

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,573

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,574

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,579

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/080,314

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/080,312

PRIOR FILING DATE: 1998-04-01

PRIOR APPLICATION NUMBER: 60/078,578

PRIOR FILING DATE: 1998-03-19

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 470

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 26

LENGTH: 1751

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (1520)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1557)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1689)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1729)

OTHER INFORMATION: n equals a,t,g, or c

FEATURE:

NAME/KEY: SITE

LOCATION: (1741)

OTHER INFORMATION: n equals a,t,g, or c

US-10-653-595-26

## Alignment Scores:

Pred. No.:	9.82e-11	Length:	1751
Score:	154.00	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-10-653-595-26 (1-1751)

QY 1 MetargarglYAlaGlyAlaAlaarglYArGAlaSerTPCySTrPAlaLeuAlaLeu 20

Db 49 ATGCCGCGGGCGCGCGCGCGCTCGGGACGCGCTTCTGTGCTGGGCGCGCTG 108

QY 21 LeutrpLeuAlaValaValaProgly 28

Db 109 CTTGGCTCGCGGTGGTTCGGGCG 132

JS-09-397-945-26	RESULT 4	
Sequence 26, Application US/09397945		
Publication No. US20030065139A1		
GENERAL INFORMATION:		
APPLICANT: Human Genome Sciences, Inc. et al.		
TITLE OF INVENTION: 95 Human secreted proteins		
FILE REFERENCE: P2027P1		
CURRENT APPLICATION NUMBER: US/09/397,945		
CURRENT FILING DATE: 1999-09-17		
PRIOR APPLICATION NUMBER: PCT/US99/05804		
PRIOR FILING DATE: 1999-03-18		
PRIOR APPLICATION NUMBER: 60/078,566		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,576		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,573		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,574		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,579		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/080,314		
PRIOR FILING DATE: 1998-04-01		
PRIOR APPLICATION NUMBER: 60/080,312		
PRIOR FILING DATE: 1998-04-01		
PRIOR APPLICATION NUMBER: 60/078,578		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,581		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,577		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/078,563		
PRIOR FILING DATE: 1998-03-19		
PRIOR APPLICATION NUMBER: 60/080,313		
PRIOR FILING DATE: 1998-04-01		
NUMBER OF SEQ ID NOS: 470		
SOFTWARE: PatentIn Ver. 2.0		
SEQ ID NO 26		
LENGTH: 1751		
TYPE: DNA		
ORGANISM: Homo sapiens		
FEATURE:		
NAME/KEY: SITE		
LOCATION: (1520)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1557)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1689)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1729)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1735)		
OTHER INFORMATION: n equals a,t,g, or c		
NAME/KEY: SITE		
LOCATION: (1741)		
OTHER INFORMATION: n equals a,t,g, or c		
JS-09-397-945-26		
Alignment Scores:		
Pred. No.:	9,82e-11	length: 1751
Score:	154.00	Matches: 28
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatch: 0
Query Match:	100.00%	Indels: 0
GB:	13	Gaps: 0
JS-10-010-050A-2_COPY_1_28 (1-28) x US-09-397-945-26 (1-1751)		

```

QY      1 MetArgArgGlyAlaGlyAlaValaArgGlyArgAlaSerTTPCySTPAlaLeuAlaLeu 20
Db      49 ATCCGGCGGGGGCGCGGGCGCGCGGTCTGGSGAGCGCGTTCTGTGCTTGCGCCTCGCGCTG 100
QY      21 LeuTriPLeuAlaValaValProGly 28
Db      109 CTTTGACTCGCGGTGTGATTCGGGAC 132

RESULT 5
US-09-122-383-13
; Sequence 13, Application US/09122383A
; Patent No. US20020042093A1
GENERAL INFORMATION:
APPLICANT: Shepard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38
CURRENT APPLICATION NUMBER: US/09/122,383A
CURRENT FILING DATE: 1998-07-24
EARLIER FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 13
LENGTH: 1038
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Degenerate nucleotide sequence encoding zsig46
OTHER INFORMATION: polypeptide of SEQ ID NO:2
FEATURE:
NAME/KEY: variation
LOCATION: (1)...(1038)
OTHER INFORMATION: n is any nucleotide
US-09-122-383-13

Alignment Scores:
Pred. No.:          9.68e-05      Length:        1038
Score:              106.00         Matches:         19
Percent Similarity: 67.86%         Conservative:    0
Best Local Similarity: 67.86%       Mismatches:     9
Query Match:        68.83%         Indels:          0
DB:                  9             Gaps:            0

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-122-383-13 (1-1038)
QY      1 MetArgArgGlyAlaGlyAlaValaArgGlyArgAlaSerTTPCySTPAlaLeuAlaLeu 20
Db      1 ATGMGNMGNGGNCNGCNGCNGCNGGNGGNGGNGGNGGNSNTGTTGTGGCCYTNNGCNYYTN 60
QY      21 LeuTriPLeuAlaValaValProGly 28
Db      61 YTMGTGYTNGCNGTNGTNCNCGN 84

RESULT 6
US-10-010-050A-13
; Sequence 13, Application US/10010050A
; Publication No. US20020173624A1
GENERAL INFORMATION:
APPLICANT: Shepard, Paul O.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN
FILE REFERENCE: 97-38C1
CURRENT APPLICATION NUMBER: US/10/010,050A
CURRENT FILING DATE: 2002-03-26
PRIOR APPLICATION NUMBER: US 09/122,383
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/053,613
PRIOR FILING DATE: 1997-07-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
```



```

: 13 Gaps: 0
-10-010-050a-2_copy_1_28 (1-28) x US-10-027-632-241716 (1-580)
      4 G1YAlaG1YAlaAlaArgG1YArgAlaSerTTPCySTTPAlaLeuAlaLeuTTPLeu 23
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      22 GGTATCGCCAGTGCAGAGAGTGGGTTTCCATGGTGTGGGAGCTCCATCTCTGTGCTT 81
      24 AlaValVal 26
      |||
      82 TGCAGGGTA 90

SHUT 10
-10-027-632-241716
Sequence 241716, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 241716
LENGTH: 580
TYPE: DNA
ORGANISM: Human
-10-027-632-241716

Alignment Scores:
ad. No.: 15 Length: 580
Percent Similarity: 64.00 Matches: 11
Best Local Similarity: 56.52% Conservative: 2
Mismatch: 47.83% Mismatches: 10
Indels: 0
Gaps: 0
-10-010-050a-2_copy_1_28 (1-28) x US-10-027-632-241716 (1-580)
      4 G1YAlaG1YAlaAlaArgG1YArgAlaSerTTPCySTTPAlaLeuAlaLeuTTPLeu 23
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      22 GGTATCGCCAGTGCAGAGAGTGGGTTTCCATGGTGTGGGAGCTCCATCTCTGTGCTT 81
      24 AlaValVal 26
      |||
      82 TGCAGGGTA 90

SHUT 11
-10-282-122A-15190
Sequence 15190, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
```

```

; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15190
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: Bordetella pertussis
; US-10-282-122A-15190

Alignment Scores:
Pred. No.: 27.9 Length: 1071
Score: 64.00 Matches: 11
Percent Similarity: 60.87% Conservative: 3
Best Local Similarity: 47.83% Mismatches: 9
Query Match: 41.56% Indels: 0
Gaps: 0
US-10-010-050a-2_copy_1_28 (1-28) x US-10-282-122A-15190 (1-1071)
QY 2 ArgArgG1YAlaG1YAlaAlaArgG1YArgAlaSerTTPCySTTPAlaLeuAlaLeu 21
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 AGGCGTGGCGGAGGCTGGCGCATATCCGGACGGACGTGGTGGGCGATGGCGGCTATG 310
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 22 TTPLeuAla 24
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 311 TGGCTTCC 319
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-10-087-192-754/C
; Sequence 754, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
```



; PRIOR FILING DATE: 2001-03-02  
 ; NUMBER OF SEQ ID NOS: 2059  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 754  
 ; LENGTH: 276276  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-10-087-192-754

Alignment Scores:  
 Pred. No.: 8.84e+03  
 Score: 276276  
 Length: 276276  
 Matches: 13  
 Percent Similarity: 52.00%  
 Conservative: 0  
 Best Local Similarity: 52.00%  
 Mismatches: 11  
 Query Match: 41.23%  
 Indels: 1  
 Gaps: 1

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-10-087-192-754 (1-276276)

Qy 4 GlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeuLeu---Trp 23  
 Db 221348 GGAGCCACGACACCCGCGGCTTCTTCAGCTGCTGCGGCTGGGGCTGGCCCTCGG 221289

Qy 23 LeuAlaValAlaPro 27  
 Db 221288 CTGTACTCGGTGCGG 221274

RESULT 13

US-10-029-517-106  
 ; Sequence 106, Application US/10029517  
 ; Publication No. US20030148969A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kenneth W. Dobie  
 ; APPLICANT: Susan J. Myers  
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF MUCIN 1, TRANSMEMBRANE EXPRESSION  
 ; FILE REFERENCE: RTS-0352  
 ; CURRENT APPLICATION NUMBER: US/10/029,517  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 107  
 ; SEQ ID NO 106  
 ; LENGTH: 772  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 US-10-029-517-106

Alignment Scores:

Pred. No.: 27  
 Score: 63.00  
 Length: 772  
 Matches: 13  
 Percent Similarity: 60.00%  
 Conservative: 2  
 Best Local Similarity: 52.00%  
 Mismatches: 8  
 Query Match: 40.91%  
 Indels: 2  
 Gaps: 1

S-10-010-050A-2\_COPY\_1\_28 (1-28) x US-10-029-517-106 (1-772)

Y 4 GlyAlaGlyAlaAlaArg-----GlyArgAlaSerTrpCysTrpAlaLeuAlaLeuLeu 21  
 b 551 GGAGCTGGGGTGGCCAGGCTGGGGCATGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 61C

Y 22 TrpLeuAlaValAla 26  
 b 611 GCGCTGGCCATTGTC 625

35UTR 14

3-10-425-114-5227/c

Sequence 5227, Application US/10425114  
 Publication No. US20040034888A1  
 GENERAL INFORMATION:

APPLICANT: Liu, Jingdong  
 APPLICANT: Zhou, Yihua  
 APPLICANT: Kovalic, David K.  
 APPLICANT: Screen, Steven E

; APPLICANT: Tabaka, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; FILE REFERENCE: Plants and Uses Thereof for Plant Improvement  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 5227  
 ; LENGTH: 1257  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: 700448276\_FU1  
 US-10-425-114-5227

Alignment Scores:

Pred. No.: 44.1  
 Score: 63.00  
 Length: 1257  
 Matches: 10  
 Percent Similarity: 66.67%  
 Conservative: 0  
 Best Local Similarity: 66.67%  
 Mismatches: 5  
 Query Match: 40.91%  
 Indels: 0  
 Gaps: 0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-10-425-114-5227 (1-1257)

Qy 2 ArgArgGlyAlaAlaArgGlyArgAlaSerTrpCysTrp 16  
 Db 881 GGAGCAGCAGCAGAGCGCGGCGGCGGCTTCTTGGGCTCGG 837

RESULT 15

US-10-354-247-1/c  
 ; Sequence 1, Application US/10354247  
 ; Publication No. US20030224393A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gudmundsson Gudmundur  
 ; TITLE OF INVENTION: GENE FOR PERIPHERAL ARTERIAL OCCLUSIVE  
 ; FILE REFERENCE: DISEASE  
 ; CURRENT APPLICATION NUMBER: US/10/354,247  
 ; CURRENT FILING DATE: 2003-01-29  
 ; PRIOR APPLICATION NUMBER: US 10/060,902  
 ; PRIOR FILING DATE: 2002-01-30  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 322101  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)-(322101)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-10-354-247-1

Alignment Scores:

Pred. No.: 1.2e+04  
 Score: 63.00  
 Length: 322101  
 Matches: 11  
 Percent Similarity: 60.00%  
 Conservative: 1  
 Best Local Similarity: 55.00%  
 Mismatches: 8  
 Query Match: 40.91%  
 Indels: 0  
 Gaps: 0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-10-354-247-1 (1-322101)

Qy 4 GlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeuLeu 23  
 Db 166506 GGTGAGGCTGACGACGAGAGGTAGTCCCATGCTTGGGAGCTCCACTTCTGTGGCTT 166447

Search completed: April 25, 2004, 06:14:49  
 Job time : 95.8056 sec

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:02:52 ; Search time 9.29444 Seconds

(without alignments)  
1671.819 Million cell updates/sec

Title: US-10-010-050a-2\_COPY\_1\_28  
Sequence: 1 MRGAGAGRAGRWALALMLAVPG 28

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=x1h  
-Q=/cg2\_1/USPTO.spool/US10010050/runat\_22042004\_113205\_27569/app\_query.fasta\_1.1372  
-DB=Issued\_Patents\_NA -OPMT=Isatap -SUFFIX=ini -MINMATCH=0.1 -LOOPCL=0  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USBR=US10010050 @CGN 1.1 140 @runat\_22042004\_113205\_27569 -NCPU=6 -ICPU=3  
-NO\_MMAP -JARBOTRY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents\_NA:\*  
1: /cg2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cg2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cg2\_6/prodata/2/ina/6A.COMB.seq:\*  
4: /cg2\_6/prodata/2/ina/6B.COMB.seq:\*  
5: /cg2\_6/prodata/2/ina/PCUTUS.COMB.seq:\*  
6: /cg2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	62.5	40.6	2067	US-08-713-928B-8	Sequence 8, Appl1
C 2	62.5	40.6	2155	US-09-191-171-4	Sequence 4, Appl1
C 3	62.5	40.6	2155	US-09-385-707-4	Sequence 4, Appl1
C 4	62.5	40.6	2155	US-09-639-696C-4	Sequence 4, Appl1
C 5	62.5	40.6	4480	US-09-191-171-7	Sequence 7, Appl1
C 6	62.5	40.6	4480	US-09-385-707-7	Sequence 7, Appl1
C 7	62.5	40.6	6200	US-09-439-923-1	Sequence 1, Appl1
C 8	62.5	40.6	6200	US-09-711-202A-1	Sequence 1, Appl1
C 9	62.5	40.6	6200	US-09-711-205A-1	Sequence 1, Appl1
C 10	62.5	40.6	6238	US-09-639-696C-6	Sequence 6, Appl1
C 11	62.5	40.6	16998	US-09-676-610B-24	Sequence 24, Appl1
C 12	62.5	40.6	197496	US-09-877-177A-10	Sequence 10, Appl1

C 13	62	40.3	365	4	US-09-621-976-11459	Sequence 11459, A
C 14	61	39.6	2502	4	US-09-069-023-2	Sequence 2, Appl1
C 15	60.5	39.3	657	4	US-09-489-039A-2986	Sequence 2986, Ap
C 16	60.5	39.3	1452	4	US-09-489-039A-2735	Sequence 2735, Ap
C 17	60	39.0	894	4	US-09-489-039A-5631	Sequence 5631, Ap
C 18	60	39.0	1113	4	US-09-489-039A-5264	Sequence 5264, Ap
C 19	60	39.0	33529	3	US-09-144-085-3	Sequence 3, Appl1
C 20	59	38.3	1017	4	US-09-023-655-738	Sequence 738, Ap
C 21	59	38.3	1731	4	US-09-489-039A-688	Sequence 688, Ap
C 22	58	37.7	1371	4	US-09-489-039A-1161	Sequence 1161, Ap
C 23	58	37.7	2100	3	US-08-938-830-2	Sequence 2, Appl1
C 24	58	37.7	2100	3	US-09-020-222-2	Sequence 2, Appl1
C 25	58	37.7	111282	4	US-09-754-250-3	Sequence 3, Appl1
C 26	57.5	37.3	1077	4	US-09-016-434-1121	Sequence 1121, Ap
C 27	57.5	37.3	37030	4	US-08-311-731A-25	Sequence 25, Appl1
C 28	57	37.0	1263	4	US-09-252-991A-11039	Sequence 11039, A
C 29	57	37.0	1656	4	US-09-252-991A-11216	Sequence 11216, A
C 30	57	37.0	2799	1	US-08-446-794A-5	Sequence 5, Appl1
C 31	57	37.0	2799	1	US-08-750-007-4	Sequence 4, Appl1
C 32	56.5	36.7	53165	4	US-09-214-808-1	Sequence 1, Appl1
C 33	56	36.4	381	4	US-09-252-991A-11793	Sequence 11793, A
C 34	56	36.4	395	4	US-09-702-705-831	Sequence 831, Ap
C 35	56	36.4	395	4	US-09-736-457-831	Sequence 831, Ap
C 36	56	36.4	395	4	US-09-614-124B-831	Sequence 831, Ap
C 37	56	36.4	395	4	US-09-671-325-831	Sequence 831, Ap
C 38	56	36.4	421	4	US-09-641-638-506	Sequence 506, Ap
C 39	56	36.4	700	1	US-08-037-579A-4	Sequence 4, Appl1
C 40	56	36.4	700	3	US-08-601-184-4	Sequence 4, Appl1
C 41	56	36.4	900	4	US-09-589-927-1	Sequence 1, Appl1
C 42	56	36.4	900	4	US-09-589-927-5	Sequence 3, Appl1
C 43	56	36.4	900	4	US-09-589-927-7	Sequence 5, Appl1
C 44	56	36.4	900	4	US-09-589-927-9	Sequence 7, Appl1
C 45	56	36.4	900	4	US-09-589-927-9	Sequence 9, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-713-928B-8/C  
Sequence 8, Application US/08713928B  
Patent No. 5929304  
GENERAL INFORMATION:  
APPLICANT: RADIN, DAVID N.  
APPLICANT: CRAMER, CAROLE L.  
APPLICANT: OISHI, KAREN K.  
APPLICANT: WEISSENBOEN, DEBORAH L.  
TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/713, 928B  
FILING DATE: 13-SEP-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/003, 737  
FILING DATE: 14-SEP-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30, 742  
REFERENCE/DOCKET NUMBER: 7956-0011-999  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2067 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
US-08-713-928B-8

Alignment Scores:  
Pred. No.: 160 Length: 2067  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 40.58% Indels: 11  
Gaps: 1

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-08-713-928B-8 (1-2067)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10  
Db 1081 CGAAGGGGTGGCGGTGCTGAGTCAAGAGCATTTGCTTCTCAGAGCGGTAGGGGA 1022

QY 11 ArgAlaSerTyrCysTrrpAlaLeuAla 19  
Db 1021 AGCGCGAGGTGGTGTGGCCAGTAGCA 995

RESULT 2  
US-09-191-171-4/c  
Sequence 4, Application US/09191171  
Patent No. 6149909  
Patent No. 6149909 6143294  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
TITLE OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/191,171  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/494,104  
FILING DATE: 23-JUN-1995  
APPLICATION NUMBER: AU PK9490/91  
FILING DATE: 14-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: AU PCT/AU92/00611  
FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/084,254  
FILING DATE: 07-JUL-1993

ATTORNEY/AGENT INFORMATION:  
NAME: Digilio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2155 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 89..2047  
US-09-191-171-4

Alignment Scores:  
Pred. No.: 167 Length: 2155  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 40.58% Indels: 11  
Gaps: 1

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-09-191-171-4 (1-2155)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10  
Db 1169 CGAAGGGGTGGCGGTGCTGAGTCAAGAGCATTTGCTTCTCAGAGCGGTAGGGGA 1110

QY 11 ArgAlaSerTyrCysTrrpAlaLeuAla 19  
Db 1109 AGCGCGAGGTGGTGTGGCCAGTAGCA 1083

RESULT 3  
US-09-385-707-4/c  
Sequence 4, Application US/09385707  
Patent No. 6238662  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
APPLICANT: ANSON, Donald S.  
APPLICANT: ORSBORN, Annette M.  
APPLICANT: NELSON, Paul V.  
APPLICANT: CLEMENTS, Peter R.  
APPLICANT: MORRIS, Charles P.  
APPLICANT: HOWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
TITLE OF INVENTION: SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/385,707  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/494,104  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Digilio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8978Z  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2155 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 89..2047  
US-09-385-707-4

Alignment Scores:  
Pred. No.: 167 Length: 2155  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 40.58% Indels: 11  
Gaps: 1  
DB: 3

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-09-385-707-4 (1-2155)

Qy 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10  
Db 1169 CGAAGGGGTGGCGGTGAGTCTCAGGAAGCATTTGCTGCTCAGGAGCGGTAGGGGA 1110

Qy 11 ArgAlaSerTyrCysTyrAlaLeuAla 19  
Db 1109 AGCGGAGGTGGTGTGGCCAGTAGCA 1083

RESULT 4  
US-09-639-696C-4/c  
Sequence 4, Application US/09639696C  
Patent No. 6524835  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
ANSON, Donald S.  
ORSBORN, Annette M.  
NELSON, Paul V.  
CLEMENTS, Peter R.  
MORRIS, Charles P.  
HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON PEABODY LLP  
STREET: 990 Stewart Avenue  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/639,696C  
FILING DATE: 16-AUG-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/191,171  
FILING DATE: 13-NOV-1998  
APPLICATION NUMBER: US 08/494,104  
FILING DATE: 23-JUN-1995

APPLICATION NUMBER: AU PK9490/91  
FILING DATE: 14-NOV-1991  
APPLICATION NUMBER: AU PCT/AU92/00611  
FILING DATE: 12-NOV-1992  
APPLICATION NUMBER: US 08/084,254  
FILING DATE: 07-JUL-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pokalsky, Ann R.  
REGISTRATION NUMBER: 34,697  
REFERENCE/DOCKET NUMBER: 2249/304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 516-832-7572  
TELEFAX: 516-832-7555  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2155 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 89..2047  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-639-696C-4

Alignment Scores:  
Pred. No.: 167 Length: 2155  
Score: 62.50 Matches: 13  
Percent Similarity: 48.28% Conservative: 1  
Best Local Similarity: 44.83% Mismatches: 4  
Query Match: 40.58% Indels: 11  
Gaps: 1  
DB: 4

US-10-010-050A-2\_COPY\_1\_28 (1-28) x US-09-639-696C-4 (1-2155)

Qy 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10  
Db 1169 CGAAGGGGTGGCGGTGAGTCTCAGGAAGCATTTGCTGCTCAGGAGCGGTAGGGGA 1110

Qy 11 ArgAlaSerTyrCysTyrAlaLeuAla 19  
Db 1109 AGCGGAGGTGGTGTGGCCAGTAGCA 1083

RESULT 5  
US-09-191-171-7/c  
Sequence 7, Application US/09191171  
Patent No. 6149909  
GENERAL INFORMATION:  
APPLICANT: SCOTT, Hamish S.  
ANSON, Donald S.  
ORSBORN, Annette M.  
NELSON, Paul V.  
CLEMENTS, Peter R.  
MORRIS, Charles P.  
HOPWOOD, John J.  
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC  
SEQUENCES ENCODING SAME  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

```
APPLICATION NUMBER: US/09/191,171
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-191-171-7

Alignment Scores:
Pred. No.: 375 Length: 4480
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 3 Gaps: 1

US-10-010-050a-2_copy_1_28 (1-28) x US-09-191-171-7 (1-4480)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
Db 2192 CGAAGGGGTCGGGTAGCTCAGAGGCGATTGTCGTCACAGAGCGCGTAGG3GA 2133

QY 11 ArgAlaserTPrCystrPaAlaLeuAla 19
Db 2132 AGCGGAGGTGTGTGGCCAGTAGCA 2106

RESULT 6
US-09-385-707-7/c
; Sequence 7, Application US/09385707
; Patent No. 6238662
; GENERAL INFORMATION:
; APPLICANT: SCOTT, Hamish S.
; APPLICANT: ANSON, Donald S.
; APPLICANT: ORSBORN, Annette M.
; APPLICANT: NELSON, Paul V.
; APPLICANT: CLEMENTS, Peter R.
; APPLICANT: MORRIS, Charles P.
; APPLICANT: HOPKINS, John J.
; TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY SCOTT MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,707
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,104
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DIGILIO, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 89782
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4480 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-385-707-7

Alignment Scores:
Pred. No.: 375 Length: 4480
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 3 Gaps: 1

US-10-010-050a-2_copy_1_28 (1-28) x US-09-385-707-7 (1-4480)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
Db 2192 CGAAGGGGTCGGGTAGCTCAGAGGCGATTGTCGTCACAGAGCGCGTAGG3GA 2133

QY 11 ArgAlaserTPrCystrPaAlaLeuAla 19
Db 2132 AGCGGAGGTGTGTGGCCAGTAGCA 2106

RESULT 7
US-09-439-923-1/c
; Sequence 1, Application US/09439923
; Patent No. 6426208
; GENERAL INFORMATION:
; APPLICANT: Emily D. Kakkis
; TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods for
; TITLE OF INVENTION: for Producing and Purifying the Same and Methods for
; TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof
; FILE REFERENCE: 08000051US00
; CURRENT APPLICATION NUMBER: US/09/439,923
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6200
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1558)...(3516)
US-09-439-923-1

Alignment Scores:
Pred. No.: 537 Length: 6200
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
```

```
test Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 4 Gaps: 1
JS-10-010-050A-2_COPY_1_28 (1-28) x US-09-439-923-1 (1-6200)

Y 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
  |||||
  2638 CGAAGGGGTGGGTGTAGCTCAGGAAGCATTTGTTGCTCAGAGCGCGTAGGGGA 2579
  |||||
  Y 11 ArgAlaSerTrpCysTrpAlaLeuAla 19
  |||||
  DB 2578 AGCGGAGGTGGTGTGGCCAGTAGCA 2552

RESULT 8
JS-09-711-202A-1/C
Sequence 1, Application US/09711202A
Patent No. 6569661
GENERAL INFORMATION:
APPLICANT: Emili D. Kakkis
TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods
TITLE OF INVENTION: for Producing and Purifying the Same and Methods for
TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof
FILE REFERENCE: 08000051US00
CURRENT APPLICATION NUMBER: US/09/711.202A
PRIOR FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: US/09/439,923
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 1
LENGTH: 6200
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1558)...(3516)
JS-09-711-202A-1

Alignment Scores:
Pred. No.: 537 Length: 6200
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 4 Gaps: 1

JS-10-010-050A-2_COPY_1_28 (1-28) x US-09-711-202A-1 (1-6200)

Y 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
  |||||
  DB 2638 CGAAGGGGTGGGTGTAGCTCAGGAAGCATTTGTTGCTCAGAGCGCGTAGGGGA 2579
  |||||
  Y 11 ArgAlaSerTrpCysTrpAlaLeuAla 19
  |||||
  DB 2578 AGCGGAGGTGGTGTGGCCAGTAGCA 2552

RESULT 9
US-09-711-205A-1/C
Sequence 1, Application US/09711205A
Patent No. 6585971
GENERAL INFORMATION:
APPLICANT: Emili D. Kakkis
TITLE OF INVENTION: Recombinant Alpha-L-Iduronidase, Methods
TITLE OF INVENTION: for Producing and Purifying the Same and Methods for
TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof
FILE REFERENCE: 08000051US00
CURRENT APPLICATION NUMBER: US/09/711.205A
PRIOR FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: US/09/439,923
PRIOR FILING DATE: 1999-11-12
```

```
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 1
LENGTH: 6200
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1558)...(3516)
US-09-711-205A-1

Alignment Scores:
Pred. No.: 537 Length: 6200
Score: 62.50 Matches: 13
Percent Similarity: 48.28% Conservative: 1
Best Local Similarity: 44.83% Mismatches: 4
Query Match: 40.58% Indels: 11
DB: 4 Gaps: 1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-711-205A-1 (1-6200)

QY 2 ArgArgGlyAla-----GlyAlaAlaArgGly 10
  |||||
  DB 2638 CGAAGGGGTGGGTGTAGCTCAGGAAGCATTTGTTGCTCAGAGCGCGTAGGGGA 2579
  |||||
  QY 11 ArgAlaSerTrpCysTrpAlaLeuAla 19
  |||||
  DB 2578 AGCGGAGGTGGTGTGGCCAGTAGCA 2552

RESULT 10
US-09-639-696C-6/C
Sequence 6, Application US/09639696C
Patent No. 6524835
GENERAL INFORMATION:
APPLICANT: SCOTT, Hamish S.
ANSON, Donald S.
ORSBORN, Annette M.
NELSON, Paul V.
CLEMENTS, Peter R.
MORRIS, Charles P.
HOPWOOD, John J.
TITLE OF INVENTION: SYNTHETIC ALPHA-L-IDURONIDASE AND GENETIC
SEQUENCES ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: NIXON PEARBODY LLP
STREET: 990 Stewart Avenue
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/639,696C
FILING DATE: 16-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/191,171
FILING DATE: 13-NOV-1998
APPLICATION NUMBER: US 08/494,104
FILING DATE: 23-JUN-1995
APPLICATION NUMBER: AU PK9490/91
FILING DATE: 14-NOV-1991
APPLICATION NUMBER: AU PCT/AU92/00611
FILING DATE: 12-NOV-1992
APPLICATION NUMBER: US 08/084,254
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pokalsky, Ann R.
```

QY 21 LeuTyrPheuAlaValProGly 28



```
DB      67539 ACCCTGCTCTCT---GCACCTGGA 67519
RESULT 12
US-09-877-177A-10/c
; Sequence 10, Application US/09877177A
; Patent No. 6582919
; GENERAL INFORMATION:
; APPLICANT: K. Danenberg
; TITLE OF INVENTION: Method of determining Epidermal Growth
; TITLE OF INVENTION: Factor Receptor and HER2-Neu Gene Expression
; TITLE OF INVENTION: and Correlation of Levels Thereof With Survival
; FILE REFERENCE: 11220/120
; CURRENT APPLICATION NUMBER: US/09/877,177A
; CURRENT FILING DATE: 2001-06-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 197496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-877-177A-10

Alignment Scores:
Pred. No.:      2.43e+04      Length:      197496
Score:          62.50         Matches:      13
Percent Similarity: 53.57%      Conservative: 2
Best Local Similarity: 46.43%      Mismatches: 12
Query Match:    40.58%         Indels:      1
DB:             4            Gaps:          1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-877-177A-10 (1-197496)
QY      1 MetAArgGlyAlaGlyAlaAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
DB      75599 ATGACACACGACGAGCGCTCTACAGAGCGCGTGCACAGTGCCTGGGGTCTGACGCC 75540
QY      21 LeuTrpLeuAlaValAlaValProGly 28
DB      75539 ACCCTGCTCTCT---GCACCTGGA 75519

RESULT 13
US-09-621-976-11459/c
; Sequence 11459, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jober, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 11459
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-11459

Alignment Scores:
Pred. No.:      26.9         Length:      365
Score:          62.00         Matches:      12
Percent Similarity: 56.00%      Conservative: 2
Best Local Similarity: 48.00%      Mismatches: 11
Query Match:    40.26%         Indels:      0
DB:             4            Gaps:          0

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-621-976-11459 (1-365)
QY      2 ArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 21
```

```
DB      352 CGCATTTGGGTATATGACCAAGAGTAAAGTCCCATGATGTTGGGACAGCTTGCCCCCTG 293
QY      22 TrpLeuAlaValVal 26
DB      292 TGGCTCTGCAGGGTG 278

RESULT 14
US-09-069-023-2/c
; Sequence 2, Application US/09069023A
; Patent No. 6348573
; GENERAL INFORMATION:
; APPLICANT: Nunez, Gabriel
; APPLICANT: Inohara, Naohiro
; APPLICANT: Koseki, Takeyoshi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
; TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS
; FILE REFERENCE: UM-03333
; CURRENT APPLICATION NUMBER: US/09/069,023A
; CURRENT FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2502
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-069-023-2

Alignment Scores:
Pred. No.:      294         Length:      2502
Score:          61.00         Matches:      14
Percent Similarity: 58.06%      Conservative: 4
Best Local Similarity: 45.16%      Mismatches: 9
Query Match:    39.61%         Indels:      4
DB:             4            Gaps:          1

US-10-010-050A-2_COPY_1_28 (1-28) x US-09-069-023-2 (1-2502)
QY      2 ArgArgGlyAlaGlyAlaAlaAlaArg-----GlyArgAlaSerTrpCysTrpAla 17
DB      307 CCCCCGGGCTCAGTAGAGCGCAGGTGCGGAGATTGTGTGTAAGGAATGTGGCGATGGCG 248
QY      18 LeuAlaLeuLeuTrpLeuAlaValValProGly 28
DB      247 CTGCAGATGGCCTCCCGCTTCATGTCCTCCGGG 215

RESULT 15
US-09-489-039A-2986
; Sequence 2986, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2986
; LENGTH: 657
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2986

Alignment Scores:
Pred. No.:      76.8         Length:      657
Score:          60.50         Matches:      13
Percent Similarity: 55.56%      Conservative: 2
Best Local Similarity: 48.15%      Mismatches: 11
Query Match:    39.29%         Indels:      1
DB:             4            Gaps:          1
```

US-10-010-050A-2\_COPY\_1\_28 (1-28) X US-09-489-039A-2986 (1-657)

[illegible]

Search completed: April 25, 2004, 05:51:38  
Job time : 46.2944 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# 3M protein - nucleic search, using frame\_plus\_p2n model

run on: April 24, 2004, 22:59:01 ; Search time 51.6833 Seconds  
(without alignments)  
2301.507 Million cell updates/sec

file: US-10-010-050A-2\_COPY\_1\_28  
perfect score: 154  
sequence: 1 MRGAGARAGASWCWALALMAVPG 28

scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

searched: 3373863 seqs, 212409941 residues

total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

MODEL=frame+g2n.model -DEV=x1h  
Q=/cgn2\_1/USPTO.spool/US10010050/runat.22042004.113203.27539/app.query.fasta\_1.1372  
-DB=N Geneseq.29Jan04 -CPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOFCU=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdt  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10010050 @CGN\_1.1 819 \_runat.22042004.113203.27539 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEODURY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

N\_Geneseq.29Jan04:\*  
1: geneseq1980s:\*  
2: geneseq1990s:\*  
3: geneseq2000s:\*  
4: geneseq2001as:\*  
5: geneseq2001bs:\*  
6: geneseq2002s:\*  
7: geneseq2003as:\*  
8: geneseq2003bs:\*  
9: geneseq2003cs:\*  
10: geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	154	100.0	697	4	AAK93424 Human CDK
2	154	100.0	697	4	AAK92158 Human CDK
3	154	100.0	1486	2	AAK02855 Human zbl
4	154	100.0	1751	2	AAK24826 Human sec
5	154	100.0	2120	4	AAK94829 Human ful
6	120	77.9	630	6	ABQ40653 Oligonuc1
7	120	77.9	630	6	ABQ40652 Oligonuc1
8	120	77.9	631	6	ABQ13702 Oligonuc1

C	9	120	77.9	631	6	ABQ13703	Abq13703 Oligonuc1
	10	106	68.8	1038	2	AAK02866	Aax02866 Human deg
	11	70	45.5	882	9	ADC08897	Adc08897 Rice DNA
	12	67	43.5	2028	2	AAX54769	Aax54769 Endothel1
	13	67	43.5	2028	3	AAA34216	Aaa34216 Human ade
	14	67	43.5	2028	3	AAF20338	Aaf20338 Human end
	15	67	43.5	2028	7	ABZ96032	Abz96032 Human end
	16	67	43.5	6225	2	AAK55273	Aax55273 Human end
	17	67	43.5	6225	3	AAA44721	Aaa44721 Human ade
	18	67	43.5	6225	3	AAF20843	Aaf20843 Human mul
	19	67	43.5	35384	3	AAF21436	Aaf21436 Human enz
	20	67	43.5	35459	7	ABZ97130	Abz97130 Human nuc
	21	67	43.5	114955	2	AAK53491	Aax53491 Human nuc
	22	67	42.9	98690	6	ABK12169	Abk12169 Human DNA
	23	66	41.9	58857	3	AAA58471	Aaa58471 Nucloctid
	24	64	41.6	1071	7	ACA27320	Aca27320 Prokaryot
	25	64	41.2	5376	6	ABK84080	Abk84080 Human CDN
	26	63.5	40.9	622	4	AAK74135	Aak74135 Human imm
	27	63	40.9	772	8	AAK57039	Aax57039 Human muc
	28	63	40.9	9289	2	AAK34655	Aax34655 Hexaploid
	29	63	40.9	322101	9	AAD58431	Aad58431 Human PAO
	30	63	40.6	256	7	ABX47735	Abx47735 Bovine ES
	31	62.5	40.6	831	4	AAH04901	Aah04901 Human CDN
	32	62.5	40.6	1872	4	AAH17524	Aah17524 Human rep
	33	62.5	40.6	2067	2	AAT71754	Aat71754 Human alp
	34	62.5	40.6	2155	6	ABL64421	AbL64421 Stomach c
	35	62.5	40.6	2155	6	ABK35560	Abk35560 Gene IDUA
	36	62.5	40.6	2155	6	ABN95897	Abn95897 Gene #239
	37	62.5	40.6	2155	6	ABK5560	Abk5560 Gene #239
	38	62.5	40.6	2155	6	ABK5560	Abk5560 Gene #239
	39	62.5	40.6	2155	6	ABK5560	Abk5560 Gene #239
	40	62.5	40.6	4475	2	AAQ45995	Aaq45995 IDUA - ex
	41	62.5	40.6	6177	6	AAD38312	Aad38312 Alternati
	42	62.5	40.6	6200	3	AAZ249493	Aaz249493 Human alp
	43	62.5	40.6	6200	6	AAD24168	Aad24168 Human alp
	44	62.5	40.6	6200	6	AAD24167	Aad24167 Human alp
	45	62.5	40.6	6200	6	AAD40485	Aad40485 Human alp

## ALIGNMENTS

RESULT 1	AAK93424	standard; CDNA; 697 BP.
ID	AAK93424	
AC	AAK93424	
XX		
DT	06-NOV-2001	(first entry)
XX		
DE	Human CDNA clone representative sequence, SEQ ID NO: 1884.	
XX		
KM	Human, full length cDNA; cDNA synthesis; oligo-capping; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1130094-A2.	
XX		
PD	05-SEP-2001.	
XX		
PF	07-JUL-2000; 2000EP-00114089.	
XX		
PR	08-JUL-1999; 99UP-00194486.	
PR	11-JAN-2000; 2000UP-00187774.	
PR	02-MAY-2000; 2000UP-00183765.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,	
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H,	
DR	WPI; 2001-524255/58.	
XX		
PT	830 Primers useful for synthesizing full length cDNA clones and their use	

PT in genetic manipulation.  
XX  
XX Example 11; SEQ ID NO 1884; 1380bp + Sequence listing; English.  
XX  
CC The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been isolated  
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have  
CC been determined. Primers for synthesizing the full length cDNA are useful  
CC for clarifying the function of the protein encoded by the cDNA. The full  
CC length clones were obtained by construction of full length enriched cDNA  
CC libraries that were synthesised by the oligo-capping method. The primers  
CC enable the production of the full length cDNA easily without any special  
CC methods. The present sequence was used as the representative sequence  
CC from a human clone which was used in homology searches to identify the  
CC clone. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in CD-ROM format directly from  
CC EPO  
XX  
SQ Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;  
XX  
XX Alignment Scores:  
XX Pred. No.: 4.41e-09 Length: 697  
XX Score: 154.00 Matches: 28  
XX Percent Similarity: 100.00% Conservative: 0  
XX Best Local Similarity: 100.00% Mismatches: 0  
XX Query Match: 100.00% Indels: 0  
XX DB: 4 Gaps: 0  
XX  
XX US-10-010-050A-2\_COPY\_1\_28 (1-28) x AAK93424 (1-697)  
XX  
XX QY 1 MetAAGAGG1YAlaG1YAla1AaArgG1YArgAlaSerTrpCytrAlaLeuAlaLeu 20  
XX DB 52 ATGCGCGCGGCGCGCGCGCGGCTCGGAGACCGGCTCTGCTGCTGCGGCGCTGCGGCGCTG 111  
XX  
XX QY 21 LeuTrpLeuAlaValValProGly 28  
XX DB 112 CTTTGCTCGCGGCTGCTTCCGGGC 135  
XX  
XX RESULT 2  
XX AAK92158  
XX ID AAK92158 standard; cDNA; 697 BP.  
XX  
XX AC AAK92158;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX DE Human cDNA 5'-end sequence, SEQ ID NO: 618.  
XX  
XX KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN EP1130094-A2.  
XX  
XX PD 05-SEP-2001.  
XX  
XX PF 07-JUL-2000; 2000EP-00114089.  
XX  
XX PR 08-JUL-1999; 99JP-00194486.  
XX  
XX PR 11-JAN-2000; 2000JP-0018774.  
XX  
XX PR 02-MAY-2000; 2000JP-00183765.  
XX  
XX PA (HELI-) HELIX RES INST.  
XX  
XX PI Ota T, Nishikawa T, Isegai T, Hayashi K, Ishii S, Kawai Y;  
XX  
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
XX DR WPI: 2001-52425/58.  
XX  
XX PT 830 Primers useful for synthesizing full length cDNA clones and their use  
XX  
XX PT in genetic manipulation.  
XX  
XX PS Claim 2; SEQ ID NO 618; 1380bp + Sequence listing; English.

XX  
XX CC The invention relates to primers for synthesizing full length cDNA  
XX CC clones. 830 cDNA molecules encoding a human protein have been isolated  
XX CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have  
XX CC been determined. Primers for synthesizing the full length cDNA are useful  
XX CC for clarifying the function of the protein encoded by the cDNA. The full  
XX CC length clones were obtained by construction of full length enriched cDNA  
XX CC libraries that were synthesised by the oligo-capping method. The primers  
XX CC enable the production of the full length cDNA easily without any special  
XX CC methods. The present sequence is the nucleotide sequence of the 5'-end of  
XX CC a cDNA provided in the invention. Note: The sequence data for this patent  
XX CC did not form part of the printed specification, but was obtained in CD-  
XX CC ROM format directly from EPO  
XX  
XX SQ Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;  
XX  
XX XX Alignment Scores:  
XX Pred. No.: 4.41e-09 Length: 697  
XX Score: 154.00 Matches: 28  
XX Percent Similarity: 100.00% Conservative: 0  
XX Best Local Similarity: 100.00% Mismatches: 0  
XX Query Match: 100.00% Indels: 0  
XX DB: 4 Gaps: 0  
XX  
XX XX US-10-010-050A-2\_COPY\_1\_28 (1-28) x AAK92158 (1-697)  
XX  
XX QY 1 MetAAGAGG1YAlaG1YAla1AaArgG1YArgAlaSerTrpCytrAlaLeuAlaLeu 20  
XX DB 52 ATGCGCGCGGCGCGCGCGCGGCTCGGAGACCGGCTCTGCTGCTGCGGCGCTGCGGCGCTG 111  
XX  
XX QY 21 LeuTrpLeuAlaValValProGly 28  
XX DB 112 CTTTGCTCGCGGCTGCTTCCGGGC 135  
XX  
XX RESULT 3  
XX AAX02855  
XX ID AAX02855 standard; DNA; 1486 BP.  
XX  
XX AC AAX02855;  
XX  
XX DT 14-MAY-1999 (first entry)  
XX  
XX DE Human zsig46 DNA.  
XX  
XX KW Secreted protein; zsig46; human; chromosome 13; thyroid; disease;  
XX  
XX KW hypothyroidism; Graves' disease; thyrotoxicosis; thyroid cancer;  
XX  
XX KW Hirschsprung's disease; neuronal ceroid-lipofusiosis; Wilson disease;  
XX  
XX KW Reiger syndrome; immunosassay; detection; anti-idiotypic antibody;  
XX  
XX KW therapy; diagnostic; ss.  
XX  
XX OS Homo sapiens.  
XX  
XX PN  
XX  
XX FH Key Location/Qualifiers  
XX  
XX FT CDS 47..1087  
XX  
XX FT /\*tag= a  
XX  
XX FT /product= "zsig46"  
XX  
XX PN W09905275-A1.  
XX  
XX PD 04-FEB-1999.  
XX  
XX PF 24-JUL-1998; 98WO-US015431.  
XX  
XX PR 24-JUL-1997; 97US-0053613P.  
XX  
XX PR (ZIMO ) ZYMOGENETICS INC.  
XX  
XX PI Shepard PO, Gilbertson DG;  
XX  
XX PI WPI: 1999-142930/12.  
XX  
XX DR P-PSDB; AAW92967.  
XX  
XX PT New secreted polypeptide, zsig46, and its fragments, related fusion

PT	proteins used for diagnosis and treatment of thyroid disorders or
PS	diseases involving genes on chromosome 13.
XX	Claim 27; Page 88-90; 101pp; English.
XX	This invention describes the isolation of a novel human secreted protein
CC	zslg46 encoded by a gene on chromosome 13 which is mainly expressed in
CC	the thyroid. This product can be used to study secretion of proteins from
CC	cells and also to treat or prevent deficient expression of zslg46, which
CC	may be associated with thyroid diseases (e.g. hypothyroidism, Graves'
CC	disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that
CC	involve genes in the same region of chromosome 13 (e.g. Hirschsprung's
CC	disease, neuronal ceroid-lipofusiosis, Wilson disease and Reiger
CC	syndrome). Antibodies and other binding proteins, are used as immunoassay
CC	reagents to detect zslg46 or cells expressing it, e.g. for assessing
CC	thyroid function to produce anti-idiotypic antibodies, for affinity
CC	purification of zslg46, to screen expression libraries, for neutralise
CC	zslg46 activity, and to deliver toxins, radioisotopes etc. for
CC	therapeutic or diagnostic purposes. Agonists of the product can be used
CC	to promote growth, differentiation and proliferation of specific cell
CC	types, e.g. for treating (extra)thyroid diseases or as additive to cell
CC	cultures
XX	
SO	Sequence 1486 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;
SO	
Alignment Scores:	
Pred. No.:	1.03e-08
Score:	154.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	2
	Gaps: 0
US-10-010-050A-2_COPY_1_28 (1-28) x AA02855 (1-1486)	
QY	1 MetArgArgGlyValAlaGlyValAlaAlaArgGlyValAlaAlaSerTyrCysTTPAlaLeuAlaLeu 20
DB	47 ATGCGGCGGAGCGCGCGGCGGCTCGGAGCACGCGCTTCTGATGCTGAGGCGCTG3CGCTG 108
QY	21 LeuTTPLeuAlaValAlaProGly 28
DB	107 CTTGGCTCGCGGTGTTCGGGC 130
RESULT 4	
AAZ24826	
ID	AAZ24826 standard; DNA; 1751 BP.
XX	
AC	AAZ24826;
XX	
DT	02-DEC-1999 (first entry)
XX	
DE	Human secreted protein gene 16 clone HMZAD77.
XX	
XX	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; disease; cancer; tumour; neurodegenerative disorder; leukaemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS	
XX	Homo sapiens.
XX	
XX	WO947540-A1.
PN	
PD	23-SEP-1999.
XX	
XX	18-MAR-1999; 99WO-US005804.
XX	
PR	19-MAR-1998; 98US-0078563P.
PR	19-MAR-1998; 98US-0078565P.
PR	19-MAR-1998; 98US-0078573P.

**Pt** 19-MAR-1998; 98US-0076574P.  
**PR** 19-MAR-1998; 98US-0076576P.  
**PR** 19-MAR-1998; 98US-0078577P.  
**PR** 19-MAR-1998; 98US-0078578P.  
**PR** 19-MAR-1998; 98US-0078579P.  
**PR** 19-MAR-1998; 98US-0078580P.  
**PR** 19-MAR-1998; 98US-0078581P.  
**PR** 19-MAR-1998; 98US-0078582P.  
**PR** 01-APR-1998; 98US-0080313P.  
**PR** 01-APR-1998; 98US-0080315P.  
**XK** PA  
XX XX  
PA (HUMA-) HUMAN GENOME SCI INC.

Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Soppet DR,  
Pi Wei Y, Andres GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;  
PI Shi Y, Moore PA;  
DR MPI; 1999-562050/47.  
XX P-PsDB; AAU41323.

New isolated human genes, useful for diagnosis and treatment of e.g.  
PT cancer, neurological disorders, immune diseases, inflammation or blood  
PS disorders.

Claim 1; Page 308; 484pp; English.

This sequence represents a nucleic acid molecule which encodes a secreted  
CC human protein. The gene number, and the clone it is derived from, are  
CC detailed in the descriptor line. The gene can be used to generate fusion  
CC proteins by linking to the gene to a human immunoglobulin Fc portion  
CC (e.g. AAZ2480) for increasing the stability of the fused protein as  
CC compared to the human protein only. The invention relates to 95 novel  
CC genes and their fragments (nucleic acid sequences: AAZ24811-Z24907; amino  
CC acid sequences AAU41308-Y41404) which are useful for preventing, treating  
CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,  
CC pathological conditions can be diagnosed by determining the amount of the  
CC new polypeptides in a sample or by determining the presence of mutations  
CC in the new polymynucleotides. Specific uses are described for each of the  
CC 95 polymynucleotides, based on which tissues they are most highly expressed  
XQ in (see AAZ24811 for described uses)

SQ Sequence 1751 BP; 477 A; 355 C; 380 G; 523 T; 0 U; 16 Other;

**Alignment Scores:**

Pred. No.:	1.24e-08	Length:	1751
Score:	154.00	Matches:	28
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
Difference:	2	Gaps:	0

US-10-010-050A-2\_COPY\_1\_28 (1-28) x AAZ24826 (1-1751)  
QY 1 Metkargatgaglaalgaialaalaaargllyaralaserttpcyrtfpalaaleualaen 20  
Db 49 ATGGGCGGGCGGGCGGGCGGCCCTTCGTATGAGTGCCTGGCCTGGCGCTG 108  
| |||||  
Qy leutrgleualvalValProgly 28  
| |||||  
Db 109 CTTTGGCTCGCGTGtggtttccgggc 132

**RESULT 5**  
**ID ID** AAA94829 standard; cDNA; 2120 BP.  
**AAK94829;**  
**AC AC** AAA94829;  
**DT DT** 06-NOV-2001 (first entry)  
**DE DE** Human full-length cDNA, SEQ ID NO: 3977.  
**KM KM** Human, full length cDNA; cDNA synthesis; oligo-capping; ss.  
**CS CS** Homo sapiens.

```

XX PN BP1130094-A2.
XX PD 05-SEP-2001.
XX PF 07-JUL-2000; 2000EP-00114089.
XX PR 08-JUL-1999; 99JJP-00194486.
XX PR 11-JAN-2000; 2000JP-00118774.
XX PR 02-MAY-2000; 2000JP-00183765.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Oseuki T, Koga H;
XX WP1: 2001-524255/58.
XX P-PSDB; AAM93870.
XX PT 830 Primers useful for synthesizing full length cDNA clones and their use
XX PT in genetic manipulation.
XX PS Claim 8; SEQ ID NO 3977; 1380bp + Sequence Listing; English.
XX PS The invention relates to primers for synthesizing full length cDNA
XX CC clones. 830 cDNA molecules encoding a human protein have been isolated
XX CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
XX CC been determined. Primers for synthesizing the full length cDNA are useful
XX CC for clarifying the function of the protein encoded by the cDNA. The full
XX CC length clones were obtained by construction of full length enriched cDNA
XX CC libraries that were synthesised by the oligo-capping method. The primers
XX CC enable the production of the full length cDNA easily without any special
XX CC methods. The present sequence is a full length human cDNA of the
XX CC invention. Note: The sequence data for this patent did not form part of
XX CC the printed specification, but was obtained in CD-ROM format directly
XX CC from EPO
XX SQ Sequence 2120 BP; 590 A; 409 C; 460 G; 661 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1.53e-08 Length: 2120
XX Score: 154.00 Matches: 28
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 4 Gaps: 0
XX
XX US-10-010-050A-2_COPY_1_28 (1-28) x AAK94829 (1-2120)
XX
XX QY 1 MetArgArgGlyIyIaGlyIaIaAaAGGlyAAAGAlaSetTPCySTPAlaLeuAlaLeu 20
XX DB 52 ATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 111
XX QY 21 LeuTPrLeuAlaIaValaIaProGly 28
XX DB 112 CTTTGCGCTCGGGGTGGTTCGGGCG 135
XX
XX RESULT 6
XX ABQ40653/c
XX ID ABQ40653 standard; DNA; 630 BP.
XX AC
XX ABQ40653;
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27244.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX KW drug; side effect; cancer; central nervous system; cardiovascular;
XX KW gastrointestinal; respiratory system; single nucleotide polymorphism;
XX KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX OS

```

```

XX MO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001MO-EP010074.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX PR 05-SEP-2000; 2000DE-01044543.
XX
XX (EPiG-) EPIGENOMICS AG.
XX PA
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX MPI, 2002-371829/40.
XX DR
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful for
XX PT diagnosis and prognosis, comprises selective hybridization of amplicons
XX FT from chemically treated DNA.
XX PS
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ4411 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
SQ Sequence 630 BP, 219 A; 255 C; 70 G; 86 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 3.67e-05 Length: 630
XX Score: 120.00 Matches: 21
XX Percent Similarity: 75.00% Conservative: 0
XX Best Local Similarity: 75.00% Mismatched: 7
XX Query Match: 77.92% Indels: 0
XX DB: Gaps: 0
XX
US-10-010-050A-2_COPY_1_28 (1-28) x ABQ40653 (1-630)
XX
QY 1 MetATCAAGCGTGYAAlGtYAAAlaAlaArgGtYARGAlAsErTPCYrTpAlaleuAlaleu 20
Db 332 ATGCCGGCGGGGGCGGGCGCGGGCTTCGAGACCGGTTTTGtGttGgtGTTTGcGctTG 273
QY 21 LeuTrPLeuAlaValVaProGly 28
Db 272 TTTTGGTTTGCgCGGTGGTTTCGGGT 249
XX
RESULT 7
ABQ40652
ID ABQ40652 standard; DNA; 630 BP.
XX
XX AC ABQ40652;
XX
XX 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 27243.

```





DB 269 TTTTGGTTCCGGGTGTTTCCGGGT 292

RESULT 9  
ABQ13703/c  
ID ABQ13703 standard; DNA; 631 BP.  
XX  
XX ABQ13703;  
XX  
XX 12-JUL-2002 (first entry)  
XX  
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 294.  
XX  
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200218632-A2.  
XX  
XX 07-MAR-2002.  
XX  
XX 01-SEP-2001; 2001WO-EP010074.  
XX  
XX 01-SEP-2000; 2000DE-01043826.  
XX  
XX 05-SEP-2000; 2000DE-01044543.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX  
XX WPI; 2002-371829/40.  
XX  
XX Determining the degree of cytosine methylation in genomic DNA, useful for  
XX diagnosis and prognosis, comprises selective hybridization of amplicons  
XX from chemically treated DNA.  
XX  
XX  
XX Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one member,  
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the  
CC degree of hybridisation to both classes is determined from the label on  
CC the amplicon. From the ratio of labels hybridised to the two classes of  
CC oligomers, the degree of methylation is calculated. The method is used:  
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs  
CC and of a wide range of diseases, e.g. cancer, disorders of the central  
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,  
CC particularly by detecting mutations or single nucleotide polymorphisms  
CC (SNP's); and (ii) for differentiation of cell or tissue types and for  
CC investigating cell differentiation. The method allows the methylation  
CC status of many C residues to be determined simultaneously. ABQ13410-  
CC ABQ54121 represent genomic DNA sequences used to illustrate the method  
CC for determining the degree of cytosine methylation described in the  
CC disclosure of the invention  
XX  
XX  
SQ Sequence 631 BP; 220 A; 256 C; 70 G; 85 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 3.68e-05 Length: 631  
Score: 120.00 Matches: 21  
Percent Similarity: 75.00% Conservative: 0  
Best Local Similarity: 75.00% Mismatches: 7  
Query Match: 77.92% Indels: 0  
Gaps: 0  
DB: 6

US-10-010-050a-2\_COPY\_1\_28 (1-28) x ABQ13703 (1-631)

QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeu 20  
DB 423 ATCGCGCGGGCGCGCGCGCGGTTCGGGACGCGGTTTGGTGGGTGGCGCTTG 364  
QY 21 LeuTrpLeuAlaValValProGly 28  
DB 363 TTTTGGTTCCGGGTGTTTCCGGGT 340

RESULT 10  
AAK02866  
ID AAK02866 standard; DNA; 1038 BP.  
XX  
XX AAK02866;  
XX  
XX 14-MAY-1999 (first entry)  
XX  
XX Human degenerate zsig46 DNA.  
XX  
XX Secreted protein; zsig46; human; chromosome 13; thyroid; disease;  
XX hypochondriasm; Graves' disease; thyrotoxicosis; thyroid cancer;  
XX Hirschsprung's disease; neuronal ceroid-lipofunctionosis; Wilson disease;  
XX Reiger syndrome; immunosassay; detection; anti-idiotypic antibody;  
XX therapy; diagnostic; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO905275-A1.  
XX  
XX 04-FEB-1999.  
XX  
XX 24-JUL-1998; 98WO-US015431.  
XX  
XX 24-JUL-1997; 97US-0053613P.  
XX  
XX (ZYMO) ZYMOGENETICS INC.  
XX  
XX Sheppard PO, Gilbertson DG;  
XX  
XX WPI; 1999-142930/12.  
XX  
XX  
XX New secreted polypeptide, zsig46, and its fragments, related fusion  
XX proteins - used for diagnosis and treatment of thyroid disorders or  
XX diseases involving genes on chromosome 13.  
XX  
XX  
XX Claim 31; Page 94-95; 101pp; English.

CC This invention describes the isolation of a novel human secreted protein,  
CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in  
CC the thyroid. This product can be used to study secretion of proteins from  
CC cells and also to treat or prevent deficient expression of zsig46, which  
CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'  
CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that  
CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's  
CC disease, neuronal ceroid-lipofunctionosis, Wilson disease and Reiger  
CC syndrome). Antibodies and other binding proteins, are used as immunoassay  
CC reagents to detect zsig46 or cells expressing it, e.g. for assessing  
CC thyroid function to produce anti-idiotypic antibodies, for affinity  
CC purification of zsig46, to screen expression libraries, to neutralise  
CC zsig46 activity, and to deliver toxins, radioisotopes etc. for  
CC therapeutic or diagnostic purposes. Agonists of the product can be used  
CC to promote growth, differentiation and proliferation of specific cell  
CC types, e.g. for treating (extra)thyroid diseases or as additive to cell  
CC cultures  
XX  
XX  
SQ Sequence 1038 BP; 189 A; 106 C; 179 G; 166 T; 0 U; 398 Other;

Alignment Scores:  
Pred. No.: 0.00277 Length: 1038  
Score: 106.00 Matches: 19  
Percent Similarity: 67.86% Conservative: 0  
Best Local Similarity: 67.86% Mismatches: 9  
Query Match: 68.83% Indels: 0  
Gaps: 0  
DB: 2



CC	pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
CC	hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, ad
CC	well as all types of cancers which may metastasize or have metastasized
CC	to the lungs, including breast and prostate cancer
XX	
SQ	Sequence 2028 BP; 8 A; 740 C; 792 G; 484 T; 0 U; 4 Other;
Alignment Scores:	
Pred. No.:	210
Score:	67.00
Percent Similarity:	54.17%
Best local Similarity:	54.17%
Query Match:	43.51%
DB:	2
Gaps:	1
US-10-010-050A-2_COPY_1_28 (1-28) x AAx54769 (1-2028)	
QY	4 G VAlAG A A A A A G G T A A A S E R T P C Y S T P A A A A A L e U T P l e u 23
Db	206 GGgGCCCGGGGGGGCGCGGGGGGACGTCTTGCTGGGGGCTGG-----GGGTGCCTGTGGCTGG 258
QY	24 AlAvalA Pro 27
Db	260 CGCGTTGCCCG 271
RESULT 13	
ID AAA34216	
ID AAA34216 standard; DNA; 2028 BP.	
XX AAA34216;	
DT 28-JUL-2000 (first entry)	
DE Human adenosine receptor related polynucleotide SEQ ID NO:1905.	
KM Human; adenosine receptor; low adenosine antisense oligonucleotide;	
KM phosphothioate; impaired respiration; inflammation; allergy;	
KM allergic disease; bronchoconstriction; inhibitor; antiinflammatory;	
KM antiallergic; antiasmatic; cytostatic; analgesic; impaired airway;	
KM lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;	
KM respiratory distress syndrome; pain; cystic fibrosis; emphysema;	
KM pulmonary hypertension; chronic obstructive pulmonary disease; COPD;	
KM cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.	
XX Homo sapiens.	
OS WO200009525-A2.	
PN 24-FEB-2000.	
PD 03-AUG-1999; 99WO-US017712.	
PF 03-AUG-1998; 98US-0095212P.	
PR (UYEC-) UNIV EAST CAROLINA.	
PA NYce JM;	
PI WEI; 2000-205971/18.	
DR New antisense oligonucleotides useful for treating e.g. pulmonary	
PT vasoconstriction, inflammation, allergies, asthma, hypertension,	
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or	
PT cancers.	
XX Disclosure, Page 504-505; 1343pp; English.	
XX The present invention describes a new composition comprising an antisense	
XX oligonucleotide (ON) with low adenosine (up to 15%), which targets	
XX nucleic acids involved in bronchoconstriction, allergies, and/or	
XX inflammation. The ON can have antiinflammatory, antiallergic,	

	CC	antiaesthematic; cytostatic and analgesic activities. The compositions are useful for the treatment of diseases associated with inflammation,
	CC	impaired airways, including lung disease and diseases whose secondary effects afflict the lungs of a subject. They can be used for treating e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma;
	CC	impeded respiration, respiratory distress syndrome, pain, cystic fibrosis, pulmonary hypertension, emphysema, chronic obstructive pulmonary disease (COPD), and cancers such as leukaemias, lymphomas, carcinomas, and cancers which may metastasise to the lungs, including breast and prostate cancer. The reduction of the adenosine content of the ONS reduces side effects. The A-containing ONS break down with the release of deoxyadenosine which activates adenosine receptors causing bronchoconstriction and inflammation. AAA32313 to AAA35312 represent nucleotide sequences given in the sequence listing from the present invention, which correspond to SEQ ID NO:1 to 2915, and then the last 18 sequences are also called SEQ ID NO:1 to 185, but the sequences differ from the previously named sequences. SEQ ID NO:11 to 1680 (AAA3232 to AAA33992) are specifically claimed ONS from the present invention. N.B. Sequences given in the disclosure of the present invention do not match up with their corresponding SEQ ID NO: sequences given in the sequence listing
XX	SEQ	Sequence 2028 BP, 8 A, 740 C, 792 G, 484 T, 0 U, 4 Other;
	Alignment Scores:	
	Pred. NO.:	210
	Score:	67.00
	Percent Similarity:	54.17%
	Best Local Similarity:	54.17%
	Query Match:	43.51%
	DB:	3
		Gaps: 1
US-10-010-050A-2_COPY_1_28 (1-28) x AAA34216 (1-2028)		
OY	4	G1VALAG1VAL1AA1AARG1VARG1ASERTTTCGTCTTAAlaLeuLauLeuTrrPleu 23
Dd	206	CGGGCCGCGGGGCCCGGGGGCCCTGTCTTGAGGCTTG-----GGTGCCTGTGGCTG 25
OY	24	AlaValAlPro 27
Dd	260	CCGGTTGCCCGG 271
RESULT 14		
AAF20338		
ID	AAF20338 standard, DNA; 2028 BP.	
XX	AAF20338;	
AC		
XX	14-MAR-2001 (first entry)	
DT		
XX	Human endothelial nitric oxide synthase polynucleotide fragment #1905.	
DE		
XX	Low adenosine antitense oligonucleotide; phosphorothioate; allergy;	
KW	human; airway disorder; bronchoconstriction; lung inflammation;	
KM	surfactant depletion; respiratory; bronchodilator; antiinflammatory;	
KM	immunosuppressive; antiaesthetic; analgesic; hypotensive; cyostatic;	
KM	respiratory obstruction; pulmonary obstruction; impeded respiration;	
KM	surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;	
KM	respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;	
KM	pulmonary hypertension; emphysema; pulmonary transplantation rejection;	
KM	chronic obstructive pulmonary disease; pulmonary infection; bronchitis;	
KX	cancer; ss.	
XX		
OS	Homo sapiens.	
FN	WO200062736-A2.	
XX		
PD	26-OCT-2000.	
XX		
PF	24-MAR-2000; 200OMO-US008020.	
XX		
PR	06-APR-1999; 99US-0127958P.	

[illegible]

KM	Human; antisense; lung dysfunction; nasal airway dysfunction;
CC	antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KM	antiallergic; hypotensive; immunosuppressive; cytostatic; gene therapy;
XX	antisense gene therapy; respiratory; lung; adenosine sensitivity;
KM	adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
XX	lung inflammation; respiratory disease; ds.
OS	Homo sapiens.
XX	
PN	WO200285308-A2.
PD	31-OCT-2002.
XX	
FE	23-APR-2002; 2002WO-US013135.
PR	24-APR-2001; 2001US-0286137P.
XX	(EPig-) EPIGENESIS PHARM INC.
PI	Nyce JW, Li Y, Sandrasegara A, Katz E, Pabalan J, Aguilar D,
XX	Miller S, Tang L, Shanabuddin S;
DR	WPI; 2003-229219/22.
PT	Pharmaceutical composition for treating ailments associated with impaired
PT	respiration, has oligo(s) antisense to specific gene(s) or its
PT	corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
XX	ubiquinone.
PS	Disclousure; SEQ ID NO 11274; 872pp; English.
XX	
CC	The invention relates to a novel pharmaceutical composition, which has a
CC	first active agent comprising an oligonucleotide antisense to the
CC	initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC	5' and 3' intion-exon junctions, or regions within 2-10 nucleotides of
CC	junctions of genes encoding a polypeptide associated with lung and/or
CC	nasal airway dysfunction and a second active agent comprising an
CC	antiinflammatory steroid and ubiquinone. A composition of the invention
CC	has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC	immunosuppressive, and cytostatic activity. The composition may have a
CC	use in antisense gene therapy. The composition is useful for treating or
CC	preventing a respiratory, lung or malignant disease or condition, also
CC	for enhancing the prophylactic or therapeutic respiratory effect of an
CC	antiinflammatory steroid in a subject, for reducing or depleting levels
CC	of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC	receptor, producing bronchodilation, increasing levels of ubiquinone or
CC	lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC	lung inflammation, lung allergies, or a respiratory disease or condition.
CC	Note: The sequence data for this patent is not represented in the printed
CC	specification, but was obtained in electronic format directly from WIPO
XX	at ftp.wipo.int/pub/published_pct_sequences
SQ	Sequence 2028 BP; 8 A; 740 C; 792 G; 484 T; 0 U; 4 Other;
Alignment Scores:	
Pred. No.:	210
Score:	67.00
Percent Similarity:	54.17%
Best Local Similarity:	54.17%
Query Match:	43.51%
DB:	7
Gaps:	1
US-10-010-050A-2_COPY_1_28 (1-28) x ABZ96832 (1-2028)	
OY	4 GYALAGLYALALALARGSLVARGALASERTTTCYGTTPALALEUALALEUTPLEU 23
Db	206 GGCGGCGGGGGGGCGGGGGGCCCGCTTTCTGTGGCCCTTG-----GGATGCCTGTGGCTG 259
OY	24 ALAVAlvalPro 27
Db	260 CCGGTTGCCCGC 271

Search completed: April 24, 2004, 23:26:32  
JOB time : 55.6833 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - nucleic search, using frame\_plus\_p2n model

Run on: April 24, 2004, 23:02:32 ; Search time 4391.8 seconds

(Without alignments)  
2352.639 Million cell updates/sec

## SUMMARIES

29: gpb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

```
File: US-10-010-050A-2
Perfect score: 1927
Sequence: 1 MRGAGAGARASMCWALAL.....IKTYERIPPIPKTKLSGL 346

Scoring table:
  BLOSUM62
  Xgapop 10.0 , Xgapext 0.5
  Ygapop 10.0 , Ygapext 0.5
  Fgapop 6.0 , Fgapext 7.0
  Delop 6.0 , Delext 7.0

Searched: 2751329 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Command line parameters:
MODEL=frame+g2n.model -DEV=x1h
Q=/cgm_1/USPTO.spool/US10010050/runat_22042004_113204_27556/app_query.faste_1.1372
DB=EST -QPMT=fastcap -SUFFIX=est -MINMATCH=0.1 -LOOPCH=0 -LOOPEXT=0
UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45
DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
USER=US10010050.@CCN 1.1.6283.@runat_22042004_113204_27556 -NCPU=6 -ICPU=3
NO_MMAP -LARGESUBSTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:
  EST:
  1: em_estba:*
  2: em_estbm:*
  3: em_estin:*
  4: em_estma:*
  5: em_estov:*
  6: em_estpl:*
  7: em_estro:*
  8: em_hic:*
  9: gpb_est1:*
  10: gpb_est2:*
  11: gpb_hic:*
  12: gpb_est3:*
  13: gpb_est4:*
  14: gpb_est5:*
  15: em_estfun:*
  16: em_estom:*
  17: em_gss_hum:*
  18: em_gss_inv:*
  19: em_gss_pln:*
  20: em_gss_vit:*
  21: em_gss_fun:*
  22: em_gss_mam:*
  23: em_gss_mus:*
  24: em_gss_pro:*
  25: em_gss_rtd:*
  26: em_gss_png:*
  27: em_gss_vit1:*
  28: gpb_gss1:*
```

Result No.	Score	Query Match	Length	DB ID	Description
1	1739	90.2	1201	9	AL546472
2	1451.5	75.3	2356	11	AK043247
3	1443.5	74.9	2433	11	AK085741
4	1429	74.2	2247	11	AK032293
5	1394	72.3	923	13	BX331615
6	1386.5	72.0	924	14	CA488543
7	1360	70.6	869	12	BI090566
8	1312	68.1	787	12	BI223533
9	1280.5	66.5	1077	13	BX342662
10	1245	64.6	1098	10	BF982158
11	1237	64.2	689	13	BU621797
12	1214	63.0	1201	9	AL571805
13	1211	62.8	931	10	BE873363
14	1207	62.6	641	14	CB215456
15	1187.5	61.6	974	13	BQ319385
16	1167	60.6	908	13	BX370641
17	1158	59.7	915	12	BI080303
18	1149.5	59.7	938	13	BQ926170
19	1128	58.5	676	12	BG820113
20	1113	57.8	755	13	BQ443453
21	1099.5	57.1	777	10	AM475730
22	1097.5	57.0	872	12	BI734549
23	1073.5	55.7	828	12	BI156421
24	1070	55.5	607	12	BM849262
25	1056	54.8	637	14	CK128221
26	1054.5	54.7	833	12	BI456284
27	1018.5	52.9	915	13	BX757396
28	1012	52.5	564	14	CB897292
29	1012	52.5	623	13	BQ417448
30	996.5	51.7	796	12	BI182240
31	989	51.3	565	9	AU280382
32	981.5	50.9	650	14	CF724176
33	969.5	50.3	867	13	BX757383
34	952	49.4	868	14	CF593022
35	951	49.4	767	9	AJ451053
36	948.5	49.2	854	12	BI082103
37	922	47.8	869	14	CF378929
38	910.5	47.2	661	13	BY734042
39	900	46.7	922	14	CA789305
40	899	46.7	607	28	AQ309525
41	898	46.6	847	13	BQ719074
42	893	46.3	720	12	BI111605
43	886	46.0	736	12	BI077452
44	881.5	45.7	706	10	BF739857
45	880	45.7	516	14	CB455600

## ALIGNMENTS

```
RESULT 1
LOCUS AL546472 1201 bp mRNA linear EST 31-MAY-2003
DEFINITION AL546472 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1030XJ01 5-PRIME, mRNA sequence.
ACCESSION AL546472
VERSION AL546472.2 GI:31268306
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1201)
```





```

JOURNAL: 10 (11), 1757-1771 (2000)
MEDLINE: 20530913
PUBMED: 11076861
REFERENCE: 4
TITLE: The RIKEN Genome Exploration Research Group Phase II Team and the
JOURNAL: PANTOM Consortium.
AUTHORS: Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
TITLE: The PANTOM Consortium and the RIKEN Genome Exploration Research
JOURNAL: Group Phase I & II Team.
AUTHORS: Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6
(Jases 1 to 2356)
JOURNAL: Aichi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carminci, P.,
AUTHORS: Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiroaka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanakawa, T.,
Kato, H., Kawa, J., Kojima, S., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kunihiro, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akihira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Miyamatsu, M. and Hayashizaki, Y.
JOURNAL: Direct Submission
AUTHORS: Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT: cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers
ZATURES: 1. 2356
SOURCE: /organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:A730075N08"
/db_xref="MGI:2408405"
/db_xref="taxon:10090"
/cclone="A730075N08"
/tissue_type="cerebellum"
/cclone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="7 days neonate"
1. 2356
/misc_feature /note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5
(CNS PROTEIN) [Homo sapiens] [SWISSPROT|O75503, evidence:
FASTA, 73.9kD, 86.7klength, match:1041]"
ALIGN:
Alignment Scores:
red. No.: 3.04e-135 Length: 2356
score: 1451.50 Matches: 261
Percent Similarity: 83.91% Conservative: 31
Percent Local Similarity: 75.00% Mismatch: 47
Query Match: 75.32% Indels: 9
3: 11 Gaps: 3
3-10-010-050A-2 (1-346) x AK043247 (1-2356)
1 MechArgArgGlyAlaGlyAlaAlaAlaArgGlyArgAlaSerTTP-----CysTTPAlaLeu 18
|||||
|||||
|||||
|||||

```

Db	55	ATGTCGGCGGGGGCCATGAG-----GGCACTGGCGCCGGCGCTGGCGTG	105
QY	19	AlaLeuLeuTrpLeuAlaValValProGlyTyrSerArgValSerGlyIleProSerArg	38
Db	106	GCGCTGCTCGGGCTGGGAGACGATTTCTGGGGCGCGCCGCCGAGCTGTGG-----	153
QY	39	ArgHisTrpProValProTyrLeuArgPheAspPheArgProLysProAspProTyrCys	58
Db	154	CAGCGTGGCGGCGCTTACAGCGCTTCTCTTCCGTCCGAAAGACAGATCCCTACTGT	213
QY	59	GlnAlaLysTyrThrPheCysProThrGlySerProIleProValMetGlnGlyAspAsp	78
Db	214	CAACCTAGATATCTTCTCTGTCCATCCGGCTCGCCATCCAGATTAAGAAAGACATAC	273
QY	79	AspIleGluValPheArgLeuGlnAlaProValTrpGluPheTyrGlyAspLeuLeu	98
Db	274	GTCACTCAGGCTTACGACTACAGCCCGCATTTGGGAATTAATATAGAGACTCTCTG	333
QY	99	GlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsn	118
Db	334	GGACACTTTAAACTTAATGACAGAGCCGTGGATTCAGAGACACTGACAGGACAGAC	393
QY	119	TyrThrMetGluTyrPyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArg	138
Db	394	TACCAATTAGTGTATGAACTTTTCCAGCTGGGCACTGTAATTTCCCACTCCGG	453
QY	139	ProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaLysPheGluGlyIle	158
Db	454	CCTGACAAAGACGCTCCCTCTGTGTATACCAAGGGGACGCTCTTTTGAAGAAATA	513
QY	159	AspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsn	178
Db	514	GATATATAAACCTGGAAAGAAAGGGACACTGCAGTCTGTGAACCATATCCGGAAC	573
QY	179	MetPheAsnGlnMetAlaLysTyrIleValLysGlnAspAsnGluTrpGlyIleTyrTyrGlu	198
Db	574	ACATTTAACAAAGTGGCCGAGTGGTATAGCAGACATGAATCGGATTTATTAAG	633
QY	199	ThrTrpAsnValLysValAspProGlnLysGlyAlaGluThrTrpPheAspSerTyrAsp	218
Db	634	ACATGGAGCGGTCCGAGCGGCCGACGAAGGGGCCGACAGTGGTTCGATCTCTAAG	693
QY	219	CysSerLysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLys	238
Db	694	TGTTCCGATTTTGTCTTAAGGACATATAGAAATTTGCTGAATTTGGAAACGAAATTCAG	753
QY	239	AsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGly	258
Db	754	AAGATAGAAACAACTATACGAAATATTTCTTACAGTGAAGAGCTATTTACTTGGGA	813
QY	259	AsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArg	278
Db	814	AATGAACATCTATTTTGGGCCCAAGGAACAAGACTCTGTGGCCATATAAAAA	873
QY	279	PheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGln	298
Db	874	TTTATATGGCCCTTCAGACCGTATTTGTCAACCAAGATTTTCTGATGAATTTCTTGAA	933
QY	299	IlePheAspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrp	318
Db	934	ATTTTATATCAGATTAATACAGACAGCTTCACTGTATTATTAATCTTGAGATATGG	993
QY	319	PheLeuProMetLysPheProPheIleLysIleThrTyrGluGlnIleProLeuProIle	338
Db	994	TTTCTTACCAATGAAGACCCCTTTGTCAAAATACATACGAAGAAACCCGTTACCTAAC	1053
QY	339	ArgAsnLysThrLeuSerGlyLeu	346
Db	1054	CGACATACAAACATTTACCGACTTG	1077

RESULT 3

AK085741

AK085741

2433 bp

mrna

linear

HTC 20-SEP-2003

DEFINITION Mus musculus 10 days lactation, adult female mammary gland cDNA, RIKEN full-length enriched library, clone: D730033P03  
PRODUCT: similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5 PROTEIN) [Homo sapiens], full insert sequence.  
ACCESSION AK085741.1 GI:26102938  
VERSION AK085741.1  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, S., Kawai, J., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsubara, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
JOURNAL MEDLINE 11076861  
PUBMED 11076861  
REFERENCE 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
JOURNAL MEDLINE 11076861  
PUBMED 11076861  
REFERENCE 6 (bases 1 to 2433)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirokane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kigawa, I., Kishikawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, C., Sakazume, N., Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-research@riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
FEATURES  
SOURCE location/Qualifiers  
1..2433  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="G57BL/6J"  
/db\_xref="FANTOM DB:D730033P03"  
/db\_xref="MGX:242937"  
/db\_xref="taxon:10090"  
/clone="D730033P03"  
/sex="female"  
/tissue\_type="mammary gland"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days lactation, adult"  
1..2433  
/note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5 PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence: PASTY, 73.9%ID, 86.7%length, match=1041)"  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,046-134 Length: 2433  
Score: 1443.50 Matches: 260  
Percent Similarity: 83.62% Conservative: 31  
Best Local Similarity: 74.71% Mismatches: 48  
Query Match: 74.91% Indels: 9  
DB: 11 Gaps: 3  
US-10-010-050a-2 (1-346) x AK085741 (1-2433)  
1 MetArgArgG1yAlaG1yAlaAlaArgG1yArgAlaSerTrp-----CysTrpAlaLeu 18  
132 ATGCTGCGGGGGGGGGCAGTGTGA-----GGCAGCTGGGGCGGGCGCTGGCTG 182  
19 AlaLeuLeuTrpLeuAlaValAlaProG1yTrpSerArgValSerG1yLeuProSerArg 38  
183 GCGCTGCTGGGCTGGCGAGCATTCGTGGCGGCGCTCCCGACGCTGG----- 230  
39 ArgHisTrpProValProTyrTrpArgPheArgPheArgProLysProAspProTyrCys 58  
231 CAACGCTGCGCGGTGCTGCAAGCGCTTCTCTTCCGTCCGAAGACAGTCCCTGCTGT 230  
59 GlnAlaLeuTyrTrpPheCysProThrGlySerProIleProValMetGlnGlyAspAsp 78  
291 CAAGCTAAGTATATCTTCTGCTCTACCGGCTCGCCATCCGATTATGAAGACATGAC 350  
79 AspIleGlnValPheArgLeuGlnAlaProValTyrGlnPheLysTyrGlyAspLeuLeu 98  
351 GTCATGAGAGTCTTACGACTACCAAGCCCGATTGGAAATTAAATGAGACCTCTCG 410  
99 GlyHisLeuLeuLysLeuMetHisAspAlaIleGlyPheArgSerThrLeuTrpGlyLysAsn 118  
411 GGCACATTTAAATTATGATGATGACGCGGCGGATTCAGAGACACACTGACAGCAAGAAC 470  
119 TyrThrMetGluTrpTyrGlyLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArg 138  
471 TACACATTGAGAGTATGATGACTTTCCAGCTGGGCACTGATTCATTTCCCACTCCGG 530  
139 ProGluMetAspAlaProPheTrpCysPheGlnGlyAlaAlaCysPhePheGlnGlyLeu 158  
531 CCTGACAGAGCCCTCTCTCTGCTGATTCACCAAGGGGAGCCCTTTTGAAGAAATA 500  
159 AspAspValHisTrpTyrGlyLeuGlnGlyTrpLeuValGlnValAlaThrIleSerGlyAsn 178  
591 GATGATTAACACTGGAAGAAACGGAACACTGCTAGCTGCTTCCACATCATCGGAAC 650  
179 MetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGlnThrGlyLeuTyrTrpGln 198  
651 ACATTTAACAAAGTGGCGAGTGGGTGAAGCAAGCAATGAATCGGATTATATGAG 710

199 ThTTpAnVallyAlaserProGluVgAlaGluThrTrpPheAspSerTyrAsp 218  
 711 ACATGAGACGGTCCGAGCCGCTCAGAGACAGGGGCGGACGAGCTTCAGTCTCAGAC 770  
 219 CysSerLySphValLeuArgThrPheAsnLyLeuAlaGluPheGlyAlaGluPheLyS 238  
 771 TGTTCGAAATTTTCTTAAAGACATATAGAAATTTGGCTGCAATTTGGACAGAAATTCAG 830  
 239 AsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGly 258  
 831 AAGTGAAGAAACAACATACAGAAATTTCTTACAGTGAAGAGCTATTTATCTGAGG 890  
 259 AsnGluThrSerValPheGlyProThrGlyAsnLySphLeuGlyLeuAlaIleLySArg 278  
 891 AATGAACAATCTATTTTGGGCGCAAGAAAGAAAGAAAGCTCTTCTTGGCAATAAAAA 950  
 279 PheTyrTyrProPheLySProHISLeuProThLySLeuPheLeuSerLeuLeuGln 298  
 951 TTTTATGGCCCCCTTCAGACCGGATTTGTCAACCAAGATTTTGTGATGAATTTCTTGA 1010  
 299 IlePheAspAlaValIleValHISLySglnPheTyrLeuPheTyrAsnPheGlyTyr 318  
 1011 ATTTTGATATCAGTGATTATACACAGACAGTCTACTGTTTATTAATCTTGATATGG 1070  
 319 PheLeuProVellySphPhePheIleLySILEThrTyrGluGluIlePheLeuProIle 338  
 1071 TTTCTCAATGAAAGAAACCCCTTTGTCAAAATTAACAACGAAAGAAACCCGTTACTACC 1130  
 339 ArgAsnLySphLeuSerGlyLeu 346  
 1131 CGACATACAAATTTTACCGACTTG 1154

SULT 4  
 .032293  
 CUS 2247 bp mRNA linear HTC 18-SEP-2003  
 FINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
 enriched library, clone:6430516P20 product:similar to  
 CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLMS PROTEIN) [Homo  
 sapiens], full insert sequence.  
 CESSION AKO32293  
 RSION AKO32293.1 GI:26328112  
 YWORDS HTC; CAP trapper.  
 URCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Carninci, P. and Hayashizaki, Y.  
 TITLE High-efficiency full-length cDNA cloning  
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
 MEDLINE 99279253  
 PUBMED 10349636

REFERENCE  
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
 MEDLINE 20499374  
 PUBMED 11042159

REFERENCE  
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
 Komno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,  
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, M.,  
 Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
 Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M.,  
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer  
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
 MEDLINE 20530913

PUBMED 11076861  
 REFERENCE 4  
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the  
 FANTOM Consortium.  
 TITLE Functional annotation of a full-length mouse cDNA collection  
 JOURNAL Nature 409, 685-690 (2001)  
 REFERENCE 5  
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research  
 Group Phase I & II Team.  
 TITLE Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 JOURNAL Nature 420, 563-573 (2002)  
 REFERENCE 6 (bases 1 to 2247)  
 AUTHORS Adachi, J., Aizawa, K., Akimura, T., Aikawa, T., Bono, H., Carninci, P.,  
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,  
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,  
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,  
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
 Tanaka, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
 Muramatsu, M. and Hayashizaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of  
 Physical and Chemical Research (RIKEN), Laboratory for Genome  
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
 Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)  
 COMMENT CDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in Riken.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL:http://genome.gsc.riken.go.jp/  
 URL:http://fantom.gsc.riken.go.jp/  
 Location/Qualifiers  
 1..2247  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:6430516P20"  
 /db\_xref="WGI:2395762"  
 /db\_xref="taxon:10090"  
 /clone="6430516P20"  
 /sex="male"  
 /tissue\_type="olfactory brain"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 <1..971  
 /note="unnamed protein product; putative  
 similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLMS  
 PROTEIN) [Homo sapiens] (SWISSPROT|075503, evidence:  
 FASTY, 73.9%ID, 86.7%length, match=1041)"  
 /codon\_start=3  
 /protein\_id="BAC27797.1"  
 /db\_xref="GI:26328113"  
 /translation="IGLATILGASPRSGORWVPVYKSPRPKTPDYCOAKYFPCPTG  
 SPFVPMKDKDLEAVTLQAPLPIEFKYGDLGKPKLMDVAGPSTITGKNTLEWEL  
 FQIGNGTDFHLPRDKSAPWCGACGFEGLDDKMKNGKLTSLVATISGNTFNKVA  
 EMTKQDNENGIYEYETWTRVAGGCGACGFESEYDCNPFRLRYKLAEGETEPKLET  
 NTYKILYGESEPIYENETISIGPKGNKTLALAIKFGPFRFYISTKDFLNMFLKIF  
 DVTIHRQYVLFNFEYWFLLPMKPEPVKITYETPTPTHTPTTDI"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 5,32e-133 Length: 2247  
 Score: 1429.00 Matches: 251

```

Percent Similarity: 86.50% Conservative: 31
Percent Local Similarity: 76.99% Mismatches: 40
Query Match: 74.16% Indels: 4
3: 11 Gaps: 1
3-10-010-050A-2 (1-346) x AK032293 (1-2247)

Y 21 LeuThrPheuAlaValAlaProGlyTrpSerArgValSerGlyLeuProSerArgHis 40
Y 3 CTGGGGCTGGGAGCAGATTCGGGGGGCTCCCGACGTCGGG-----CAAGCC 50
Y 41 TrpProValProTyrLysArgPheAspPheArgProLysProAspProTyrCysGlnAla 60
Y 51 TGCGCGGATGCATCAAGACACTTCCTCTTCGCTCCGAAGACAGATCCCTCACTGTCAAGCT 110
Y 61 LysTyrThrPheCysProThrGlySerProIleProValMetGluLysPheAspIle 80
Y 111 AAGATACCTTCCTGCTCTACCGGCTGCCCATCCAGTTATGAAGACATAGACGTATC 170
Y 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAspLeuGlnHis 100
Y 171 GAGGTCTTACGACTACACAGCCCGCATTTGGGATTTAAATATGAAACCTCTCGGGACAC 230
Y 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
Y 231 TTATAACTTATGTCATGACGGCGGTGGGATTCAGAGACACTGACAGGCAAGAACTTCACAA 290
Y 121 MetGluTrpTyrGluLeuPheGlnLeuGlnValAsnCysThrPheProHisLeuArgProGlu 140
Y 291 ATTGAGGGATGATGAACCTTTCGAGCTGGGCACTGATCACTTCCCACTCGGGCTGAC 350
Y 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAsp 160
Y 351 AAGAGCCCTCCCTCTGCTGTAACCAAGGGGAGCCGCTTTTATGAAGATATGATGAT 410
Y 161 ValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
Y 411 AAACACTGGAAAGGAAAAAGGAGACCTGTCACTGCTGTAACCACTATCCGAAACACATT 470
Y 181 AsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyLeuTyrTrpGluThrTrp 200
Y 471 AACAAAGTGGCGGATGGGTGAAGCAAGACATGAATCTGGATTTATATATGACATGG 530
Y 201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSer 220
Y 531 ACGGTCCGACGCGGCCACAGACAAAGGGGCCACAGACGTGTTCAGATCTTACGACTGTGG 590
Y 221 LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240
Y 591 AATTGTGCTTAAAGACATATTAAGAAATTGGCTGATATTGGACAGATTCAGAAAGATTA 650
Y 241 GluThrAsnTyrThrArgGluPheLeuTyrSerGlyGluProThrTyrLeuGlnAsnGlu 260
Y 651 GAACCAAACTATACGAAAAATATTTCTTTACGTGGAGAGGCCCTTATACCTGGGAAATGAA 710
Y 261 ThrSerValPheGlyLysProThrGlyAsnLysPheLeuGlnLeuAlaIleLysArgPheTyr 280
Y 711 ACATTTATTTTGGGCCCAAGAAACAGACTCTTGTCTTGGCCATTAATAAATTTTAT 770
Y 281 TyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePhe 300
Y 771 GGCCCTTCAGACCGTATTTGTCAACCAAAATTTTCTGATGATATTTCTGAAAATTTT 830
Y 301 AspAlaValIleValHisLysGlnPheLysTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320
Y 831 GATACAGGATTTATACAGACAGCTTCACTGTGTTTAACTTGTGATATGGTTTCTA 890
Y 321 ProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsn 340
Y 891 CCAATGAACCCCTTTGTCAAAATTAATATCGAAGAAACCCGTTACCTACCGGACAT 950
Y 341 LysThrLeuSerGlyLeu 346
|||||
|||
:::
|||

```

Db	951	ACAACTTACGACTTG	968
RESULT 5	EX331615	923 bp	mRNA linear EST 01-MAY-2003
LOCUS	EX331615		
DEFINITION	EX331615 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens		
ACCESSION	EX331615		
VERSION	EX331615.1	GI:30310073	
KEYWORDS	EST.		
ORGANISM	Homo sapiens (human)		
SOURCE	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1, W.B., Gruber, C., Jesse, J. and Polayes, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7238.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CSODB009AB01QPL&cluster=7238.f. Contact : Peng Liang Email : filiang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Paradey Avenue Genoscope sequence ID : CSODB009AB01QPL.		
FEATURES	Location/Qualifiers		
SOURCE	1..923		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="CSODB009YC01"		
	/tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"		
	/clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"		
	/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with NotI and cloned into the NotI and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."		
ORIGIN			
Alignment Scores:			
Pred. No.:	5.55e-130	Length:	923
Score:	1394.00	Matches:	251
Best Local Similarity:	94.36%	Conservative:	0
Best Local Similarity:	94.36%	Mismatches:	15
Query Match:	72.34%	Indels:	0
DB:	13	Gaps:	0
US-10-010-050A-2 (1-346) x EX331615 (1-923)			
QY	1	MetAArgAgGyAlaGyAlaAlaAlaGyAlaArgAlaSerTyrCysTrrAlaLeuAlaLeu	20
Db	118	ATGGGGCGGGGCGGGCGGGCGGGCTCGGGAGACGGCGCTTCCTCGGTGCTGGGCGCTG	177
QY	21	LeuTrpLeuAlaAlaValAlaProGlyTrrSerArgValSerGlyIleProSerArgArgHis	40
Db	178	CTTGGCTCGGGGGGCTTCGGGGCTTCGGGGCTTCGGGGATCCCTCCGGGGCGAC	237
QY	41	TrrProValProTyrGlyArgPheAspPheArgPrrGlyProAspPrrGlyCysGlnAla	60
Db	238	TGGCGCGGCGCTTACAGCGCTTACCTTCGTCACAAACCTGATCCCTTATGTCAGCT	297
QY	61	LysTyrTrpPheCysProThrGlySerProIleProValMetGluGlyAspAspAspIle	80
Db	298	AAGTACTCTTCTGTCACACCTGGCTACCTTACCCAGATTATGAGAGGATATGACATT	357
QY	81	GluValPheArgLeuGlnAlaProValTrrGluPheLysTyrGlyAspLeuLeuGlyHis	100
Db	358	GAAATTTTCATTACAGCCCGCCCACTATGGAATTAAATATGAGAGACCTCCTGGGACAC	417

101 LeuLysellIewechisAspalaalleglypheaXgSerThreuthrGlylyAsenYrThr 120  
 418 TTCAAAATTAATGATGNTGNCATTGATTCAGAGTACATTAACCTGGAAAGAACTANACA 477  
 121 MetGluTrpYrGluLeuPheGlnLeuGlyAsnCySerThreProHisLeuArgProGlu 140  
 478 ATGGATGATTAATACCTTTCCAACTGGGAANTGTANATTCCCNCTCCGNNCTGNA 537  
 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyLeaAsp 160  
 538 ATGGATGSSCCTTTCTGCGKGTATCAAGGGGSGTCCGCTTTTGGGGAAATGRTGAT 597  
 161 ValHisTrpYrGluLeuGlnGlyTrpLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180  
 598 GTTCACCTGGAAGAAATGGACATTAAGTCAAGTACCACTATACAGAAACATGTTTC 657  
 181 AspGluMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleYrYrGluThrTrp 200  
 658 AACCAATGGCAAGTGGTGGAAACAGACATGAACAGAAATTTATTAATGACATGG 717  
 201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerYrAspCysSer 220  
 718 AATGTAAACCCAGCCCAAGAAAGGGGCGAGACATGTTGATTCCTACGACTGTCC 777  
 221 LysPheValLeuArgTrpPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240  
 778 AATTTGTGTTAAGACCTTTAACAAGTTGCTGAATTTGACAGAGTTCAAGAACATA 837  
 241 GluThrAsnYrThrArgIlePheLeuYrSerGlyGluProThrYrLeuGlyAsnGlu 260  
 838 GAACCAACTATACAGAAATATTTCTTACAGTGGAGAACCTTACTTCGGAAATGAA 897  
 261 ThrSerValPheGlyPro 266  
 898 ACATCTTTTTGGGCCA 915

SULT 6  
 488543  
 CUS  
 FINITION  
 CA488543 924 bp mRNA linear EST 14-NOV-2002  
 AGENCOURT 10808884 MAPCL Homo sapiens cDNA clone IMAGE:6720309 5',  
 mRNA sequence.  
 CESSION  
 KSION  
 CA488543.1 GI:24950702  
 EST.  
 Homo sapiens (human)  
 DRCE  
 YWORDS  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 1 (bases 1 to 924)  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: gsa@b-remail.nih.gov  
 Tissue Procurement: Kristi A. Egland, Ira Pastan  
 cDNA Library Preparation: Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMNL at:  
 http://image.llnl.gov  
 Plate: LMNL4279 row: j column: 21  
 High quality sequence stop: 637.  
 Location/Qualifiers  
 1. .924  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6720309"  
 /cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231,  
 hTERT-HMEL, LNCaP"  
 /lab\_host="EMD10B"

ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,17e-129 Length: 924  
 Score: 1386.50 Matches: 258  
 Percent Similarity: 88.63% Conservative: 7  
 Best Local Similarity: 86.29% Mismatches: 21  
 Query Match: 71.95% Indels: 13  
 DB: 14 Gaps: 3

US-10-010-050A-2 (1-346) x CA488543 (1-924)

/clone lib="MAPCL"  
 /note="Vector: PCMV-SPOrt6; Site 1: EcoRV; Site 2: Not I;  
 Subtracted with brain, liver, lung, kidney and muscle.  
 Directionally cloned. Priming method: oligo-dT. Average  
 insert size: 1800 bp. Library amplification: 26,000 fold.  
 Kristi A. Egland, James V. Vincent, Robert Strausberg,  
 Bunkook Lee & Ira Pastan. Discovery of new breast  
 cancer genes encoding membrane and secreted proteins.  
 Manuscript submitted."

QY 1 MetArgArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20  
 Db 16 ATGTGGCGGGCGGGCGGGCGGGCTCGGGACGGCTTCTGCTGGCTGGCGCTGGCGCTG 75  
 QY 21 LeuTrpLeuAlaValValProGlyTrpSerArgValSerGlyIleProSerArgArgHis 40  
 Db 76 CTTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 135  
 QY 41 TrpProValProTrpYrLysArgPheAspPheArgProLysProAspProYrYrGlnAla 60  
 Db 136 TGCGCGGTGGCCCAAGAGGTTTGAATTCCTCCGCAAAACCTGATCTTATGCAAGCT 195  
 QY 61 LysTrpThrPheCysProThrGlySerProIleProValMetGluGlyAspAspIle 80  
 Db 196 AAGTATACCTTTCTGTCCAACTGGCTCACTATCCAGTTATGAGGAGGTATGATGACATT 255  
 QY 81 GluValPheArgGlnGlnAlaProValTrpGluPheLysTrpGlyAspLeuLeuGlyHis 100  
 Db 256 GAAGTTTTCGATTAACAGCCCAAGTATGGAATTTAATATGAGACCTCTGGGACAC 315  
 QY 101 LeuLysellIewechisAspalaalleglypheaXgSerThreuthrGlylyAsenYrThr 120  
 Db 316 TTGAATAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375  
 QY 121 MetGluTrpYrGluLeuPheGlnLeuGlyAsnCySerThreProHisLeuArgProGlu 140  
 Db 376 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435  
 QY 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyLeaAsp 160  
 Db 436 ATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 495  
 QY 161 ValHisTrpYrGluLeuGlnGlyTrpLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180  
 Db 496 GTTCACCTGGAAGAAATGGACATTAAGTCAAGTACCACTATACAGAAACATGTTTC 555  
 QY 181 AsnGluMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleYrYrGluThrTrp 200  
 Db 556 AACCAATGGCAAGTGGTGGAAACAGACATGAACAGAAATTTATTAATGACATGG 615  
 QY 201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerYrAspCysSer 220  
 Db 616 AATGTAAACCCAGCCCAAGAAAGGGGCGAGACATGTTGATTCCTACGACTGTTC 675  
 QY 221 LysPheValLeuArgTrpPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240  
 Db 676 AATTTGTGTTAAGACCTTTAACAAGTTGCTGAATTTGACAGAGTTCAAGAACATA 735  
 QY 241 GluThrAsnYrThrArg-IlePheLeuYrSerGlyGluPro---ThrTrpLeuGly 259  
 Db 736 GAACCAACTATACAGAAATATTTCTTACAGTGGAGAAACCTTACTTATTCGGGGA 795

[illegible]

Accession	Version	Keywords	Organism	Reference	Authors	Title	Journal	Comment
Db								
Qy	81							
Db	292							
Qy	101							
Db	352							
Qy	121							
Db	412							
Qy	141							
Db	472							
Qy	161							
Db	532							
Qy	181							
Db	592							
Qy	201							
Db	651							
Qy	220							
Db	711							
Qy	240							
Db	771							
Qy	259							
Db	831							
RESULT 8	BI223533							
LOCUS	60294192.F1							
DEFINITION	NIH_MGC_12 Homo sapiens CDNA clone IMAGE:510593 5',							
ACCESSION	BI223533							
VERSION	BI223533							
KEYWORDS	EST.							
SOURCE	GI:14676977							
ORGANISM	Homo sapiens (human)							
REFERENCE	Homo sapiens							
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
COMMENT	1 (bases 1 to 787)							
	NIH-MGC http://mgs.nci.nih.gov/							
	National Institutes of Health, Mammalian Gene Collection (MGC)							
	Unpublished (1999)							
	Contact: Robert Strausberg, Ph.D.							
	Email: cgabs-remail.nih.gov							
	Tissue Procurement: ATCC							
	CDNA Library Preparation: Life Technologies, Inc.							
	DNA Sequencing by: Incyte Genomics, Inc.							
	Clone distribution: MGC clone distribution information can be							
	found through the I.M.A.G.E. Consortium/BLN at:							
	http://image.llnl.gov							
	Plate: L14M11253 row: n column: 06							
	High quality sequence stop: 762.							
FEATURES	location/Qualifiers							
SOURCE	1..787							
	/organism="Homo sapiens"							







355 TGAAGTTTTCGATTACAGCCCGAGTATGGGAATTTAAATATGAGACCTCTGGAGCA 414  
 100 sLeuylselliEweChisaspAlailegIlyPhehysSerThleuthrGlylyAsnlyrTh 120  
 415 CTTGAAAAATATGATGATGAGCCAGTATTCAGAAATCAATTAATCTGCAAGAACTACAC 474  
 120 MetGluTrpYrGlyLeuPheGlnLeuGlyAsnCyseThrPheProhIstLeuArgProG 140  
 475 AATGGAATGATATGAACTTTTCCAACTTGGCAACTGACATTTCCCATCTCCGACTGA 534  
 140 uMetAspAlaIProhPheTrpCyseAsnGlnIyAlaIaCysePhePheGlnGlyIleAspAs 160  
 535 AAMGAGATGCCCTTCTGCTGCTATTCARGGCGCTGCTTTTAAAGGGAATGTATGA 594  
 160 PValHISTrplysGlyLeuGlnIleValGlnValAlaThrIleSerGlyAsnMetPh 180  
 595 TGTCACCTGAGAGAGAAATGGACATTAATTCARAGCAATATTCARAGAAACATGTT 654  
 180 eAsnGlnMetAlaIySerTrpValIyGlnAspAsnGlyuthrGlyIleTyTrpGlnThr 200  
 655 CAACCAATGGCAAAAGTGGGAAACAGRCATGAAACAGRAATTTWTATTTAGACATK 714  
 200 PAsnValIyAlaIaSerProGlnIyGlyAlaGluThrTrpPheAspSerTyAspCyse 220  
 715 GAATGCTMAAACARCCCAAAAAAGGGGAGAGACATGTTTATTTCTACGACTGTC 774  
 220 rlyPhePheAlaIleuArgThrPheAsnIyLeuAlaGluPheGlyAlaGluPheIyAsnI 240  
 775 CAAMTTTGTATTAAGACCTTTAACAAATTTGCTGAATTTGACAGAGITCAAPAAACAW 834  
 240 eGluThrAsnIyTrpThrArgIlePheLeuTySerGlyGluProThrTyTrpLeuGlyAsnG 260  
 835 AAAAAACAATATACAAAATATTTCTTTTAAAGTGGARACT-ACTTATCTKGGGAATGA 893  
 260 uThSerValPheGlyProhngIyAsnIyThrIleuGlyLeuAlaIlelyAsnGlyPheTy 280  
 894 AACATCTGTTTGGGAGACARGAACMAAA---CTCTTGTCTTTCGCTTAAAAAAGATTW 950  
 280 rTyTrpPheIyProhIstLeuProhIyGluPheIleuSerLeuGlnIlePh 300  
 951 TTACCCCTTCARCCCMWTGGCAMAHAATAATKTTT-----TWAGKGYTTGAAMTTT 1004  
 300 eAspAlaValIleValIstIyGlnPheTyLeu 311  
 1005 TTKBATATATTTTGWMAAATAKWTTTTTTT 1038

ESULT 10 1098 bp mRNA linear EST 23-JAN-2001  
 F982158  
 CCUS 602308926P1 NIH\_MGC\_88 Homo sapiens cDNA clone IMAGE:4400298 5',  
 EPINITIATION mRNA sequence.  
 CCESSION BF982158  
 ERSION BF982158.1 GI:12384970  
 EYWORDS  
 OURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 1098)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabds-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1AM10105 row: 9 column: 19  
 High quality sequence stop: 693.

FEATURES  
 source  
 Location/Qualifiers  
 1..1098  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4400298"  
 /tissue="duodenal adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 88"  
 /note="Organ: small intestine; Vector: pCMV-SPORT6;  
 Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally;  
 oligo-dT primed. Average insert size 1.767 kb. Library  
 enriched for full-length clones and constructed by Life  
 Technologies. Note: this is a NIH\_MGC library."

ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.61e-115 Length: 1098  
 Score: 1245.00 Matches: 240  
 Percent Similarity: 88.13% Conservative: 5  
 Best Local Similarity: 86.33% Mismatches: 30  
 Query Match: 64.61% Indels: 6  
 DB: 10 Gaps: 2

US-10-010-050a-2 (1-346) x BF982158 (1-1098)

QY 1 MetArgArgGlyAlaGlyAlaAlaArgAlaSerTrpCystrpAlaLeuAlaLeu 20  
 Db 35 ATGCGCGGGGGCGGGCGGGCGGCTCGGGAGCGGCTTCGTGCTGCGCCGCGCTG 94  
 QY 21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyIleProSerArgArgHis 40  
 Db 95 CTTGGCTCGCGGGGTTCGGCGGCTCGGGGTCTCGGGCAATCCCTCCCGCGCGCAC 154  
 QY 41 TrpProValProTyTrpArgPheAspPheArgProIyAspAspProTyTrpCysGlnAla 60  
 Db 155 TGGCGGGGTCTCAAGACCGCTTGACTTCGTCGCAAACTATCTTATGTCAGCT 214  
 QY 61 TyTrpThrPheCyseProhngIySerProIleProValMetGlnIyAspAspAspIle 80  
 Db 215 AAGTATCTTCTGTCACATCGGCTCACTATCCAGTTATGAGAGGATGATGACAT 274  
 QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheIyTrpGlyAspLeuGlyHis 100  
 Db 275 GAATTTTTCATTAACAAGCCCGAGTGGAAATTAATTAATGAGACCTCTGGAGAC 334  
 QY 101 LeuylselliEweChisaspAlailegIlyPhehysSerThleuthrGlylyAsnlyrTh 120  
 Db 335 TTGAAATATGATGATGAGCCAGTATTCAGAAATCAATTAATCTGCAAGAACTACAC 394  
 QY 121 MetGluTrpYrGlyLeuPheGlnLeuGlyAsnCystrPheProhIstLeuArgProGlu 140  
 Db 395 ATGGAATGATGAACTTTCCAACTTGGCAACTGTACATTTCCCATCTCCGACTGAA 454  
 QY 141 MetAspAlaIProhPheTrpCyseAsnGlnIyAlaIaCysePhePheGlnGlyIleAspAs 160  
 Db 455 ATGATGCCCTTCTGCTGCTATTCARGGCGCTGCTTTTAAAGGGAATGTATGA 514  
 QY 161 PValHISTrplysGlyLeuGlnIleValGlnValAlaThrIleSerGlyAsnMetPhe 180  
 Db 515 GTTCACTGAGAGACACTCGGACATTAAGTACATTAATATGAGAAACATGTTCC 574  
 QY 181 AsnGlnMetAlaIySerTrpValIyGlnAspAsnGlyuthrGlyIleTyTrpGlnThr 200  
 Db 575 AACCAATGGCAAAAGTGGGAAACAGCAATGAAACAGAAATTTATCTGAGACCTGG 634  
 QY 201 AsnValIyAlaIaSerProGlnIyGlyAlaGluThrTrpPheAspSerTyAspCyser 220  
 Db 635 CCGTCACAGCGCAG-CCAGAACAGCGGGCAGACCTGGTTGATTCCTACGACTGTTCC 693  
 QY 221 LysPheValIleArgThrPheAsnIyLeuAlaGluPheGlyAlaGluPheIyAsnIle 240  
 Db 694 AATTT-GTGTACGAGACTT-AACAAGTTGGCTGATTTGGAGCAAGATCCAGAAC-- 748

241 GlnThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrIleuGlyAsnGlu 260  
 749 ----TGAACCATATCCCGAATATTCCTCACCGGGGAAACTGATTTGGCGCATGTCAC 805  
 261 ThrSerValPheGlyProThrGlyAsnLysThrIleuGlyLeuAlaIleLysArg 278  
 806 ACTGGTGTGGGAGC---ACCGGGAACCGCCCTCGATTCTTCTCCAAAGACA 856

BU621797 689 bp mRNA linear EST 23-SEP-2002  
 US-10-010-050a-2 (1-346) x BU621797 (1-689)  
 UT-H-FL1-bgd-1-07-0-UT.s1 NCI CGAP FL1 Homo sapiens cDNA clone  
 UT-H-FL1-bgd-1-07-0-UT 3', mRNA sequence.  
 BU621797  
 BU621797.1 GI:23288012  
 EST.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 689)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: James Martin  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 Seq primer: M13 FORWARD  
 POLYA=Yes

Location/Qualifiers  
 1. 689  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-H-FL1-bgd-1-07-0-UT"  
 /tissue\_type="Cell lines"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_id="NCI CGAP FL1"  
 /note="Organ: Chondrosarcoma; Vector: pRTT3-Pac  
 (Pharmacia) with a modified polylinker; Site 1: EcoR I;  
 Site 2: Not I; NCI CGAP FL1 is a normalized cDNA library  
 derived from a pool of mRNA obtained from 4 cell lines  
 from grade III chondrosarcoma tissues. The library was  
 constructed according to Bonaldo, Lennon and Soares,  
 Genome Research, 6:791-806, 1996. First strand cDNA  
 synthesis was primed with an oligo-dT primer containing a  
 Not I site. Double stranded cDNA was ligated to an EcoR I  
 adaptor, digested with Not I, and cloned directionally  
 into pRTT3-Pac vector. The oligonucleotide used to prime  
 the synthesis of first-strand cDNA contains a library tag  
 sequence that is located between the Not I site and the  
 (dT)18 tail. The sequence tag for this library is  
 GAGGTCGGTG. The cell lines were provided by Dr. James  
 Martin from the University of Iowa.  
 TAG\_TISUB=Human Chondrosarcoma Grade 3 cell line mix  
 TAG\_LIB=UI-H-FL1  
 TAG\_SEQ=GAGTCGGTG"

Alignment Scores:  
 ed. No.: 2.63e-114 Length: 689  
 ore: 1237.00 Matches: 222  
 Percent Similarity: 100.00% Conservative: 3  
 Net Local Similarity: 98.67% Mismatches: 0  
 Very Match: 64.19% Indels: 0

DB: 13 Gaps: 0  
 US-10-010-050a-2 (1-346) x BU621797 (1-689)

QY 109 GlnPheArgSerThrIleuThrGlyAsnTyrThrMetGluTrpTyrGluLeuPheGln 128  
 Db 688 GGATTCAAGAGTACATTAAGTGGCAAGACTACACAAATGATGATGATGATGATGATGAT 629  
 QY 129 LeuGlyAsnTyrThrPheProHISLeuArgProGluMetAspAlaProPheTyrAsn 148  
 Db 628 CTGGCACTGTACATTTCCCACTCCCACTCGAATGATGATGATGATGATGATGATGAT 569  
 QY 149 GlnGlyAlaAlaCysPhePheGluGlyIleAspAspValHisTrpIleGlyAsnGlyThr 168  
 Db 568 CAGGCGCTGCTGCTGCTTTTGAAGGAATGATGATGATGATGATGATGATGATGATGAT 509  
 QY 169 LeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTyrValLys 188  
 Db 508 TTAGTTCAAGTACCACTATATCAGGAAACATGTTCAACCAATGCGCAAGTGGTGAAA 449  
 QY 189 GlnAspAsnGlnThrGlyIleTyrTyrGlnThrTrpAsnValLysAlaSerProGluLys 208  
 Db 448 CAGGACAAATGAAACAGAAATTTTGTGACACATGAAATGAAATGAAATGAAATGAAATG 389  
 QY 209 GlnAlaGlnThrTrpPheAspSerTyrAspCysSerIlePheValLeuArgThrPheAsn 228  
 Db 388 GGGGCAAGACATGCTTTGATTCCTGACATGCTTCCAAATTTGTGTAAGACCTTTAAC 329  
 QY 229 LysLeuAlaGlnPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 248  
 Db 328 AAGTTGGCTGAATTTGGAGAGGAGTTCAAGAACATGAAACCAACTATACAGAAATTTT 269  
 QY 249 LeuTyrSerGlyGluProThrTyrIleuGlyAsnGlnThrSerValPheGlyProThrGly 268  
 Db 268 CTTTACGTTGAGAAACCTACTTATTTGGGAAATGAAACATCTGTTTGGGCAACAGGA 209  
 QY 269 AsnLysThrIleuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHISLeuPro 288  
 Db 208 AACCAAGCTCTTGTTAGCCATTAAGAAATTTTATTAACCTTCAACCAACCATTTGCA 149  
 QY 289 ThrIleGlnPheLeuLeuSerIleLeuGlnIlePheAspAlaValIleValHisLysGln 308  
 Db 148 ACTTAAGAAATTTGTTGAGTCTCTTGCAAATTTTATGACAGATGATGACAAACAG 89  
 QY 309 PheTyrIleuPheTyrAsnPheGluTyrTrpPheLeuProMetLysPheProPheIleLys 328  
 Db 88 TTCTATTGTTTATTAATTTGATGATGTTTATTAACCTAAGAAATTCCTTTATTTAAA 29  
 QY 329 IleThrTyrGluGlu 333  
 Db 28 ATTAACATTTGAAAGAA 14

RESULT 12  
 AL571805/c 1201 bp mRNA linear EST 31-MAY-2003  
 LOCUS AL571805 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 DEFINITION clone CSOD1030Y01 3-PRIME, mRNA sequence.  
 ACCESSION AL571805  
 VERSION AL571805.2 GI:31293196  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1201)  
 I.L.W.B., Gruber,C., Jessee,J. and Polayes,D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On Feb 16, 2001 this sequence version replaced gi:12929467.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7238.f for more information about this cluster, see

http://www.genoscope.cns.fr/  
cgl-bin/cluster.cgi?seq=CSODI030CE01NP1&cluster=7238.f. Contact :  
Feng Liang Email : fliang@life.com URL :  
http://fulltngen.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSODI030CE01NP1.

# FEATURES

source

Location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODI030Y701"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

## ALIGN

### Alignment Scores:

red. No.: 1.15e-111 Length: 1201  
score: 1214.00 Matches: 230  
Percent Similarity: 96.64% Conservative: 7  
Best Local Similarity: 96.64% Mismatches: 3  
Jery Match: 63.00% Indels: 0  
Gaps: 0

3-10-010-050a-2 (1-346) x AL571805 (1-1201)

110 PheargSerThrLeuThrGlyAsnValThrMetGluTrpTyrGluLeuPheGlnLeu 129  
1067 TTCAGAGTACATTAATGCGCAATGTCACAAATGGAATGGTATGCA -TTTTCACACT 1010  
130 GlysAsnCythrPheProHisLeuArgProGluMetAspAlaProPheTrpCysAsnGln 149  
1009 GGCACTGATACATTTCCCATCTCCGACCTGAATGATGCCCTTCTGTGTATACAA 950  
150 GYALALALACyPhePheGluGly-11eAspPheValHisTrpPheGluAsnGlyThre 163  
949 GGGGCGCCGCTCTTTTGGAGAAATGATGATGTCCTGAGAGAAATGGGCATT 893  
169 uValGlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaIleTrpValIleG 183  
889 AGTTCAAGTAGCACTATATCAGGAACATGTTCAACCAAAATGGCAAGTGGGTGAACA 833  
189 naAspAsnGluThrGlyIleTyrTyrGluThrTrpAsnValIleAspSerProGluLysG 203  
829 GGCATATGAAACAGAAATTTATATGACATGAAATGTAAAGCCAGCCAGAAAAGG 773  
209 yAlaGluThrTrpPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsn 229  
769 GGGAGAGACATGATGATCTTACACATGTTCCAAATTTGTATAGACCTTAAACA 710  
229 sLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePheLe 249  
709 GTTGGCTGAATTTGGACACAGTTCAAGACATAGAAACCACTATACAGAAATATTTCT 580  
249 uTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAs 269  
649 TTACAGTGAAGAACCTTACTATCTGGGAATGAACAATCTGTTTGGGGCAACAGGAA 590  
269 uTyrThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuProTh 289  
589 CAAGACCTCTGTTTACCCATTAACATTTTATACCCCTTCAACACCACTTTGCCAAC 530  
289 rLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLysGlnP 309  
529 TAAAGATTTCTGTAGTCTCTTGCAAATTTTGAAGCAGTGTGTCACAAACAGTT 470  
309 eTyrLeuPheTyrAsnPheGluTyrTrpPheLeuProMetLysPheProPheIleLys 329

Db 469 CTATTTGTTATATTTTGAATTTGGTTTTCCTATGAAATTCCTTTATTAAT 410  
Cy 329 eThrTyrGluGluIleProLeuProIleArgAsnLysThrLeuSerGlyLeu 346  
Db 409 AACATATGAAGAAATCCCTTACTATCAGAAACAAACACTCTCTGTGTTA 358

RESULT 13  
LOCUS BE873363 931 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601450425P1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3854119 5',  
mRNA sequence.

ACCESSION BE873363  
VERSION BE873363.1 GI:10322139  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@b-remail.nih.gov  
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at:  
http://image.llnl.gov  
Plate: ILM9579 row: b column: 08  
High quality sequence stop: 662.

## FEATURES

source

Location/Qualifiers  
1..931  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3854119"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_65"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

## ORIGIN

Alignment Scores:  
red. No.: 1.65e-111 Length: 931  
score: 1211.00 Matches: 232  
Percent Similarity: 93.57% Conservative: 1  
Best Local Similarity: 93.17% Mismatches: 14  
Query Match: 62.84% Indels: 4  
Gaps: 1  
DB: 10

US-10-010-050a-2 (1-346) x BE873363 (1-931)

77 AspAspAspIleGluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAsp 96  
2 GATGATGACATTTGAAGTTTTCATTACAGCCCGATGGAATTTAAATATGAGAC 61  
97 LeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGly 116  
62 CTCCTGGACACTTGAAATTAATGATGATGCCATTTGATTCAGAAATCAATTAATCTGAC 121  
117 LysAsnTyrThrMetGluTrpTyrGluLeuPheGlnLeuGlyAsnCythrPheProHis 136  
122 AAGAACTACACATGGAATGATATGATCACTTTCCAACTTGGCACTGTACATTTCCCAT 181  
137 LeuArgProGluMetAspAlaProPheTrpCysAsnGlnIleValAlaCysPhePheGlu 156

```

182 CTCGACCTGAAATGATGCCCTTCTGCTGTAATCAAGCGCTGCTTTTGGAG 241
157 GYIleAspValIh:stRPySgluAsnGlyThreValGlnValAlaThrIleSer 176
242 GGAATTGATGATGTTCACTGGAAAGAAAATGGACATTAGTTCAAGTAGCACTATATCA 301
177 GYAsnMetPheAsnGlnMetAlaIleValThrValGlyGlnAspAsnGluThrGlyIleTyr 196
302 GGAACACATGTTCAACCAATGGCAAGGGGTGAACGACCAATGAAACAGAAATTTAT 361
197 TYrGluThrTPAsnValIleAlaSerProGluIuysGlyAlaGluThrTPheAspSer 216
362 TATAGACATGAAATGTAAGCCAGCCCAAGAAAGGGGCGACAGACATGTTGATTCC 421
217 TYrAspCySerIysPheValIleuArgThrPheAsnIleuAlaGluPheGlyAlaGlu 236
422 TACACATGTTCCAAATTTGTGTAAAGCCTTTAAACAAGTTGGCTGATTTGGAGCAGAG 481
237 PheIysAsnIleGluThrAsnTyrThrArgIlePheIuysGlyIuPProThrTyr 256
482 TTCAAGAACATGAAACCACTATACAGAAATTTCTTACAGTGAACCTTAT 541
257 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnIysThrLeuGlyIleuAlaIle 276
542 CTGGAAATGAAACATCTGTTTGGGCCACAGAAACAAAGACTCTTGTTAGCCATA 601
277 LysArgPheIysTyrProPheIysProIleuSleu-ProThrIysGluPheIuysSerIe 296
602 AAAAGATTATTAACCTCTCAAGCATTGGCCAACTAAAGAAATTT--CTGTAGT 658
296 uLeuGlnIlePheAspAlaValIleValIleIleGlyGlnPheIysTyrAsnPheGly 316
659 CTGCNCAATTTGATGCGGTGATTTGCCCCCAGCGCTCTATTGCT-TATTAACCTTGCA 717
316 uTYrTPheIuysProMetIysPhe 324
718 -TATTGCTTTTACCTATGAATTC 741

```

**RESULT 14**  
**3215456**  
**NCUS**  
**3FINITION**  
**5'**, mRNA sequence.  
**NCI** np06007.y1 NICHHS\_Ut1 Homo sapiens cDNA clone IMAGE:5937013  
**3CESSION**  
**3RSTION**  
**3YWORDS**  
**EST.**  
**ORGANISM**  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 641)  
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contract: Robert Strausberg, Ph.D.  
 Email: cgapdb-remail.nih.gov  
 cDNA library Preparation:  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium/LINL  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC)  
 Clone distribution: NCI-CCGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
 info@image.lnl.gov  
 Plate: L1M13163 row: E column: 14  
 Seq primer: M13R21 reverse primer (ABI).  
 Location/Qualifiers  
 1. 641  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5937013"  
 /sex="female"

```

/tissue_type="normal endometrium, late proliferative
phase, cycle day 13"
/lab_host="DH10B (TI-resistant)"
/clone_lib="NICHHS_Ut1"
/notes="Organ: uterus; Vector: pCMV-SPORT6.1.cdbb (ResGen,
Invitrogen Corporation); Site 1: NotI; Site 2: EcoRV;
Cloned unidirectionally from microquantity amounts of mRNA
from normal endometrial tissue (late proliferative phase,
cycle day 13). Average insert size 1.9 kb. Library
constructed by ResGen (Invitrogen Corporation)."
```

**ORIGIN**  
**Alignment Scores:**  
 Pred. No.: 2,54e-111 Length: 641  
 Score: 1207.00 Matches: 213  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 62.64% Indels: 0  
 DB: 14 Gaps: 0  
 US-10-010-050A-2 (1-346) x CB215456 (1-641)

```

QY 18 LeuAlaIeuIeuThrPheAsnValAlaValProGlyTyrPheArgValIleSerGlyIleProSer 37
Db 3 CTGGCGCTGCTTTGGCTCGGGGTGTTCCGGGCTGATCCGGGATCTCGGGCATCCCTCC 62
QY 38 ArgArgIstRrProValIProIyIysArgPheAspPheArgProIysProAspProIy 57
Db 63 CGGGCGACATCGCGCGGCGCCCTCAACCGCTTGTACCTCCGTCMAAACCGATCCCTAT 122
QY 58 CyGlnAlaIysTyrTPhePheCySProThrGlySerProIleProValMetGlyIleGlyAsp 77
Db 123 TGTAACTAAGTAACTATTTGTCACACTGGCTCACTTATCCAGTTATGAGAGGAT 182
QY 78 AspAspIleGluValPheArgLeuGlnAlaProValIrrGluPheIysTyrGlyAspIeu 97
Db 183 GATGACATTGAAGTTTTCGATTCAAGCCCAAGTATGGAAATTTAAATAGAGACCTC 242
QY 98 LeuGlyIstIeuIysIleMetIleAspAlaIleGlyPheAspSerThrLeuThrGlyIys 117
Db 243 CTGGGACACTTGAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302
QY 118 AsrTyrThrMetGluTTPYrGluIeuPheGlnIleuGlyAsnCySThrPheProIstIeu 137
Db 303 AACTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 362
QY 138 ArgProIuMetAspAlaProPheTTPCyAsnGlnGlyAlaAlaCySPhPheGluGly 157
Db 363 CGACCTGAAATGATGCCCTTTCTGCTGTAATCAAGCGCTGCTGCTTTTGGAGGGA 422
QY 158 ILeAspAspValIh:stRPySgluAsnGlyThreValGlnValAlaThrIleSerGly 177
Db 423 ATTATGATGTTCACTGGAAAGAAAATGGACATTAGTTCAAGTAGCACTATATCAAGGA 482
QY 178 AsnMetPheAsnGlnMetAlaIleTTPValIysGlnAspAsnGluThrGlyIleTyrTyr 197
Db 483 AACATGTTCAACCAATGGCAAGCTGGTGAACAGACATGAAACAGAAATTTAT 542
QY 198 GluThrTPAsnValIleAlaSerProGluIuysGlyAlaGluThrTPheAspSerTyr 217
Db 543 GAGACATGAAATGTAAGCCAGCCCAAGAAAGGGGCGACAGACATGTTGATTCCTAC 602
QY 218 AspCySerIysPheValIleuArgThrPheAsnIleu 230
Db 603 GACTGTTCCAAATTTGTGTTAAGACCTTTAAACAAGTTG 641

```

**RESULT 15**  
**BQ919385**  
**LOCUS**  
**DEFINITION**  
 AGNCOCURT\_8802804 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6382620  
 5', mRNA sequence.  
**ACCESSION**  
 BQ919385  
 BQ919385.1 GI:22334083

**WORDS** EST.  
**FORCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
**REFERENCE** 1 (bases 1 to 974)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
**AUTHORS** National Institutes of Health, Mammalian Gene Collection (MGC)  
**TITLE** Unpublished (1999)  
**JOURNAL** Contact: Robert Strausberg, Ph.D.  
**COMMENT** Email: cgapbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNM at:  
<http://image.llnl.gov>  
 Plate: L1C2574 row: 1 column: 13  
 High quality sequence stop: 634.

# **FEATURES** source

1. 974  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6382620"  
 /tissue\_type="cardioma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_11b="NIH\_MGC\_40"  
 /note="Organ: prostate; Vector: pOTB7; Site: 1: XhoI;  
 Site: 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCAACGAG(G). Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library."

**Alignment Scores:**  
 ted. No.: 4.12e-109 Length: 974  
 core: 1187.50 Matches: 228  
 Percent Similarity: 78.77% Conservative: 28  
 Percent Similarity: 70.15% Mismatches: 47  
 Percent Similarity: 61.62% Indels: 22  
 Gaps: 6

S-10-010-050a-2 (1-346) x BQ19385 (1-974)

Y 4 GYALAGLYALALAGLY-----ArgAlaserTp-----CysTpAlaleu 18  
 b 15 GGAATCGGATGCTGCGCGGGCCATGTGAGCGCACTGGCGCGCGCTCGCGCTG 74  
 Y 19 AIAleuLeuTpLeuAlaValAlProGlyTpSerArgValSerGlyLeuProSerArg 38  
 b 75 GCGCTCTCGGCTGCGCGCAATTCGCGCGCTGCCGACGCTGCG----- 122  
 Y 39 ArgHisTpProValProTyrTyrArgPheAspPheArgProLysProAspProTyrCys 58  
 b 123 CAACGCTGCGCGCTGCGCGCAAGCGCTCTCTTCCGTCGGAAGACAGATCCCTACTGT 182  
 Y 59 GlnAlaLeuTyrThrPheCysProThrGlySerProIleProValMetGluGlyAspAsp 78  
 b 183 CAAGCTAAGTATATCTTCTGCTTACCGGCTGCCATCCAGTTATGAAGACATGAC 242  
 Y 79 AspIleGluValPheArgLeuGlnAlaProValTyrGluPheArgTyrGlyAspLeu 98  
 b 243 GTCATGAGAGTCTTACGACTACAGCCCGATTTGGAAATTAATATGAGACCTCTCG 302  
 Y 99 GlyHisLeuLeuTyrIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsn 113  
 b 303 GGACACTTAACTTATGATGATGCGCGCTGCGATTCAGAGCACACTGACGCGCAAGAAC 362

QY 119 TyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArg 138  
 Db 363 TACCAATGAGTGGTATGAACTTTTCAGCTGGCAACTGTACATTTCCCACTCCG 422  
 QY 139 ProGluMetAspIleProPheTyrCysAsnGlnGlyAlaIleAspPheGlnGlyIle 158  
 Db 423 CCGTACAGAGCGCTCCCTTCTGTGTGACCAAGGCGAGCTGCTTTTGGAAAGATA 482  
 QY 159 AspAspValHisTyrPylGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsn 178  
 Db 483 GATGATTAACCTCGAAGGAAAGAGGACACTGTACGTGTCACCATATTCGGAAAC 542  
 QY 179 MetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluThrGlyIleTyrTyrGlu 198  
 Db 543 ACATTTAAACAAAGCGCGGAGTGGTGAAGCAGCAATGAACTGGATTTATTATGAG 602  
 QY 199 ThrTyrAsnValLysAlaSerProGluLysGlyAlaGluThrTyrPheAspSerTyrAsp 218  
 Db 603 ACATGACGCTCGAGCGCGCGCGCGAGCAAGAGGCGCGAGCTGCTTGAATCTTACAC 662  
 QY 219 CysSerLysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLys 238  
 Db 663 TGTTCGAATTTGTCTTAAAGACATATTAAGAAATTTGCTGAATTTGAAACAGATTCAAG 722  
 QY 239 AsnIleGluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGly 258  
 Db 723 AAGATAGAAACAACTATACGAAATTTCTTACAGTGAAGGCTATTACTGGA 782  
 QY 259 AsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle-LysA 278  
 Db 783 AATGAACATTTATTTTGGCCCAAGGAAACAGACTTCTGCTTGGCCATATAAAA 842  
 QY 278 TspPheTyrTyrProPheLysProHisLeu---ProThrLysGlu-PheLeuLeuSerLeu 296  
 Db 843 AATTTTATGCGCCCTTCCAGACCGAATTTGTTCACCAAGATTTCTCTGATGAATTT 902  
 QY 297 LeuGln---IlePheAspAlaVal-----IleValHisLys 307  
 Db 903 CTGAAATATTTTGTATACGCTGATTAATACCCCGAGGCTTCTTAATCTGTTTAAA 962  
 QY 308 GlnPheTyr 310  
 Db 963 ACCTTTTGG 971

Search completed: April 25, 2004, 05:45:56  
 Job time : 4408.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - nucleic search, using frame\_plus\_pzn model

on: April 24, 2004, 23:04:07 ; Search time 689.597 Seconds  
(without alignments)  
2262.168 Million cell updates/sec

file: US-10-010-050A-2

effect score: 1.927  
quence: 1 MRGAGAGAGAGACAAAL.....IKITVEPIPIPKNTLSGL 346

ring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Zgapop 6.0, Zgapext 7.0  
Delop 6.0, Delext 7.0

arched: 2907579 segs, 2254313464 residues

total number of hits satisfying chosen parameters: 5815158

minimum DB seq length: 0

minimum DB seq length: 2000000000

at-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

mand line parameters:

DEL=frame+ pzn.model -DEV=xlh  
B=cg2.1/USPTO.spool/US10010050/rnat.22042004.113205.27584/app.query.fasta\_1.1372  
B=Published Applications NA -OFMT=fastap -SUFFIX=trmb -MINMATCH=0.1  
DOPE=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62  
RANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
HR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
KALEN=2000000000 -USER=US10010050 @CGN 1 1 776 @rnat.22042004.113205.27584  
CPU=6 -ICPU=3 -NO MAP -LARGESUBSTR -NEG SCORES=0 -WAIT -DSBLOCK=100  
JNGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRESH=1 -XGAPOP=10 -XGAPEXT=0.5  
XAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

tabase: Published Applications NA:

1: /cg2.6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cg2.6/prodata/2/pubpna/US06\_PCT\_NEW\_PUB.seq:\*  
3: /cg2.6/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cg2.6/prodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cg2.6/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cg2.6/prodata/2/pubpna/US07\_PUBCOMB.seq:\*  
7: /cg2.6/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cg2.6/prodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cg2.6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cg2.6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cg2.6/prodata/2/pubpna/US09\_PUBCOMB.seq:\*  
12: /cg2.6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cg2.6/prodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
14: /cg2.6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cg2.6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
16: /cg2.6/prodata/2/pubpna/US10\_PUBCOMB.seq:\*  
17: /cg2.6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
18: /cg2.6/prodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
19: /cg2.6/prodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Query	Match	Length	ID	Description
-------	-------	--------	----	-------------

1	1927	100.0	1486	9	US-09-122-383-1	Sequence 1, Appli
2	1927	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appli
3	1916	99.4	1751	13	US-10-653-595-26	Sequence 26, Appl
4	1916	99.4	1751	13	US-09-397-945-26	Sequence 26, Appl
5	1631	84.6	1038	9	US-09-122-383-13	Sequence 13, Appl
6	1631	84.6	1038	14	US-10-010-050A-13	Sequence 13, Appl
7	886	46.0	494	9	US-09-864-761-22592	Sequence 2592, A
8	727	37.7	506	14	US-10-040-739-825	Sequence 825, App
9	684.5	35.5	474	9	US-09-864-761-5832	Sequence 5832, Ap
10	472	24.5	497	16	US-10-264-049-1859	Sequence 1859, Ap
11	359.5	18.7	351	9	US-09-867-701-2579	Sequence 2579, Ap
12	241	12.5	473	9	US-09-954-456-1363	Sequence 1363, Ap
13	121	6.3	60	10	US-09-908-975-10339	Sequence 10339, A
14	112	5.8	2667	15	US-10-198-846-9753	Sequence 9753, Ap
15	108.5	5.6	3530	16	US-10-369-493-33507	Sequence 33507, A
16	100.5	5.2	1827	13	US-10-425-114-21585	Sequence 21585, A
17	97.5	5.1	750	16	US-10-369-493-32308	Sequence 32308, A
18	97.5	5.1	986	16	US-10-369-493-36401	Sequence 36401, A
19	96	5.0	870	16	US-10-369-493-26859	Sequence 26859, A
20	96	5.0	3035	16	US-10-108-260A-1791	Sequence 1791, Ap
21	96	5.0	3580	10	US-09-822-846-21	Sequence 159, App
22	96	5.0	3580	10	US-09-374-046A-159	Sequence 159, App
23	96	5.0	3580	13	US-10-616-263-159	Sequence 127, App
24	94.5	4.9	14070	10	US-09-870-759-127	Sequence 127, App
25	94.5	4.9	14070	10	US-09-751-708A-127	Sequence 127, App
26	94.5	4.9	27189	13	US-10-087-192-1828	Sequence 1828, Ap
27	94	4.9	4170	13	US-10-262-839-91	Sequence 22365, A
28	94	4.9	3717	13	US-10-262-839-91	Sequence 91, Appl
29	94	4.9	30013	9	US-09-764-877-3297	Sequence 3297, Ap
30	94	4.9	30013	16	US-10-242-515-3297	Sequence 3297, Ap
31	93	4.8	1826	15	US-10-028-245-1	Sequence 1, Appli
32	93	4.8	4683	9	US-10-114-170-22	Sequence 22, Appl
33	93	4.8	4683	9	US-09-764-868-1494	Sequence 1494, Ap
34	93	4.8	39178	17	US-10-210-589-11	Sequence 11, Appl
35	92.5	4.8	1612	14	US-10-003-392-9	Sequence 9, Appli
36	92.5	4.8	2115	14	US-10-003-392-9	Sequence 19, Appl
37	92	4.8	1468	13	US-10-424-599-120013	Sequence 120013, A
38	92	4.8	1815	13	US-10-282-122A-23914	Sequence 23914, A
39	92	4.8	3719	13	US-10-147-493-49	Sequence 49, Appl
40	92	4.8	3719	13	US-10-145-127-49	Sequence 49, Appl
41	92	4.8	3719	13	US-10-160-503-49	Sequence 49, Appl
42	92	4.8	3719	13	US-10-143-118-49	Sequence 49, Appl
43	92	4.8	3719	13	US-10-144-993-49	Sequence 49, Appl
44	92	4.8	3719	13	US-10-158-787-49	Sequence 49, Appl
45	92	4.8	3719	13	US-10-140-024-49	Sequence 49, Appl

## ALIGNMENTS

RESULT 1  
US-09-122-383-1  
: Sequence 1, Application US/09122383A  
: Patent No. US2002042093A1  
: GENERAL INFORMATION:  
: APPLICANT: Sheppard, Paul O.  
: APPLICANT: Gilbertson, Debra G.  
: TITLE OF INVENTION: SECRETED PROTEINS ENCODED BY HUMAN  
: FILE REFERENCE: 97-38  
: CURRENT APPLICATION NUMBER: US/09/122,383A  
: EARLIER APPLICATION NUMBER: 1998-07-24  
: EARLIER FILING DATE: 1997-07-24  
: NUMBER OF SEQ ID NOS: 19  
: SOFTWARE: FastSeq for Windows Version 3.0  
: SEQ ID NO 1  
: LENGTH: 1486  
: TYPE: DNA  
: ORGANISM: Homo sapien  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: (47)...(1084)







```

b 407 ATGAGATGATGAACTTTCCAACTGGCACTGACATTTCCCACTCCGACCTGAA 466
b 141 MetAPalaPProPheTTPCySaSnGnGlyAlaAlaCyPhePheGluGlyIleAspAsp 160
b 467 ATGGATGCCCTTTTGGGTATATCAAGGCGCTGCTGTTTTTGAAGGAATTGATGAT 526
b 161 VALHSTPLPGLYGLuSnGlyThrlLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180
b 527 GTTCACTGGAGAAAGAAATGGGACATTAGTTCAAGTAGCAACTATATCAGAAACAATGTT 586
b 181 AaSnGlnMetAlaLysTrpValLysGlnAspAsnGlnThrGlyIleTyrTrpGlnThrTrp 200
b 587 AACCAAAATGGCAAGGAGTGAAGCAAGCAATGAAACGAAATTATATATGACATGG 646
b 201 AaenValLysAlaSerProGluLysGlyAlaGlnThrTrpPheAspSerTyrAspCysSer 220
b 647 AATGTAAAGCCAGCCCAAGAAAGGGGCGAGACATGTTGATTTCCACACATGTTCC 706
b 221 LysPheValLysArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240
b 707 AAATTTGTATGAAGCCTTTTAAACAAGTTGGCTGAATTTGGAGCAGAGTTCAAGAACATA 765
b 241 GluThrAsnTyrThrArgLlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGln 260
b 767 GAACCACTATACAAAGAAATATTTCTTACAGTGAAGAACTTACTATCTGGAAATGAA 826
b 261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
b 827 ACATCTGTTTGGGCGCAACAGAAACAAGACTCTTGTTGSCATTAAGATTTAT 885
b 281 TyrProPheLysProHleuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePhe 300
b 887 TACCCCTTCAAAACCAATTTGCCAATAAGAAATTTCTGTTAAGTCTCTGCAAAATTTT 945
b 301 AspAlaValIleValHleLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu 320
b 947 GATGAGTATTTGTGCAACAACGTTCTATTTGTTTAAATTTGAATATGGTTTAA 1006
b 321 PrometLysPheProPheIleLysIleThrTyrGluGlnIleProLeuProIleArgAsn 340
b 1007 CCTATGAATTCCTCTTTTATTAATAACATATGAAGAAATCCCTTATCTATCGAATAC 1056
b 341 LysThrLeuSerGlyLeu 346
b 1067 AAAACACTCTCTGGTTTA 1084

```

```

PRIOR APPLICATION NUMBER: 60/080,312
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/078,578
PRIOR FILING DATE: 1998-03-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-10-653-595-26
Alignment Scores:
Pred. No.: 3,73e-230 Length: 1751
Score: 1916.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.43% Indels: 0
Gaps: 0
DB: 13
US-10-010-050A-2 (1-346) x US-10-653-595-26 (1-1751)
QY 1 MetAArgGlyAlaGlyAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
Db 49 ATGGCGGGGGGGCGGGGGCGGGCTCGGGGACGCGCTTCTGCTGCGGCCCTGGCGCTG 108
QY 21 LeuTrpLeuAlaValAlaProGlyTyrPheArgValSerGlyIleProSerArgArgHis 40
Db 109 CTTTGGCTCGCGGGTTCGGGGCTGGTCCCGGGCTGCGGCATCCCTCCGCGGCCAC 168
QY 41 TrpProValProTyrLysArgPheAspPheArgPheArgProLysProAspProTyrCysGlnAla 60
Db 169 TGGCGGGTCCCTACAGGCGCTTATGACTTCCTGCAAAACCTGATCTTATGTCAAGCT 228
QY 61 LysTyrThrPheCysProThrGlySerProIlePheProValMetGlnGlyAspAspAspIle 80
Db 229 AAGTATACCTTTGTGTCGAACCTGCTACCTATCCAGTTATGAGAGGTGATGATGACAT 288
QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAspLeuLeuGlyHis 100
Db 289 GAAGTTTTCGATTTCAGAGCCCAAGATGGGAATTAATATAGAGACCTCTGGAGAC 348
QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120
Db 349 TTGAATAATATGCAAGATGCAATGGATATCAAGATCAATACCTGCGCAAGACTACACA 408
QY 121 MetGluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHleuArgProGlu 140

```

```

409 ATGAAATGATGTAACCTTTCCAACTTGCAACTGATCAATTCCTCCGACTGAA 463
141 MetAspAlaIapropheTrpCysaengInglYAlaAlaCysPhepheGluYlleaPaP 163
469 ATGGAATGCCCCCTTTGGTGTATCAAGGCGCTGCTTTTGAAGGAATGATGAT 523
161 ValHisTrpIysGluansGlyThrLeuValGlnValAlaThrIleSerGlyanMetPhe 183
529 GTTCACTGGAAGAAATGCGACATTAGTCAAGTGAACAATATCAGGAACATGTTTC 583
181 AsnGlnMetAlaIysTrpValIlyGlnAspAsnGlnThrGlyIleYrtyrGluThrTrp 203
589 AACCAATGCGAAAGTGGTGAACACGACAAATGAAACGGAATTTATATGACATGG 643
201 AsnValIlyAlaSerProGluIysGlyAlaGluThrTrpPheAspSerYrtyrAspCysSer 220
649 AATGTTAAAGCCGACCCGAAAGGGGGGACAGACATGTTTATTCCTACGACTGTTCC 708
221 LysPheValIleuArGThrPheAsnIlyLeuAlaGluPheGlyAlaGluPheIysAsnIle 240
709 AATTTGTGTAAAGACCTTTACAAAGTTGGCTGAATTTGACAGACAGTTCAAGACATA 768
241 GluThrAsnTrpThrArGliePheLeuYrtyrSerGlyGluProThrYrtyrLeuGlyanGlu 260
769 GAAMCCAACTATACAGAAATATTTCTTTACAGTGAACATACTATCTGGGAAATGAA 828
261 ThrSerValPheGlyProThrGlyAsnIlyThrLeuGlyLeuAlaIleIysArGpHeYr 280
829 ACATCTGTTTTGGGCAACAGAAACAGACACTCTTGCTTACCTATTAATAAATTTAT 888
281 TyrProPheIysProHisIleuProThrIlyGluPheLeuIleuSerLeuIleuGlnIlePhe 300
889 TACCCCTTCAACACACTTGGCCAACTAAAGAAATTTCTGTGAGTCTCTGCAAAATTTT 948
301 AspAlaValIleValHisIlyGlnPheYrtyrLeuPheYrtyrAsnIleuYrtyrPheLeu 320
949 GATGCAATGATTTGGCAACAAAGTTCTATTTGTTATATTTGAAATATGATTTTAA 1008
321 ProMetIysPheProPheIleIysIleThrYrtyrGluGluIleProLeuProIleArGAsn 340
1009 CCAATGAATTTCCCTTTATTAATAATTAACATATGAGAAATCCCTTACCTATCAAGAAC 1068
341 LysThrLeuSerGlyLeu 346
1069 AAAACACTCTCTGCTTTA 1086

```

```

PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,581
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,577
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/078,563
PRIOR FILING DATE: 1998-03-19
PRIOR APPLICATION NUMBER: 60/080,313
PRIOR FILING DATE: 1998-04-01
NUMBER OF SEQ ID NOS: 470
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 26
LENGTH: 1751
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1520)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-09-397-945-26
Alignment Scores:
Pred. No.: 3,73e-230 Length: 1751
Score: 1916.00 Matches: 344
Percent Similarity: 99.42% Conservative: 0
Best Local Similarity: 99.42% Mismatches: 2
Query Match: 99.43% Indels: 0
Gaps: 0
US-10-010-050A-2 (1-346) x US-09-397-945-26 (1-1751)
QY 1 MetArgArgGlyValAlaGlyAlaAlaArgGlyAlaArgAlaSerTrpCysTrpAlaLeuAlaLeu 20
DB 49 ATGCGGCGGGGCGCGGCGGCTCGGGGACGGGCTTCTGTGCTGGGCGCTGCGGCTG 108
QY 21 LeuTrpLeuAlaValAlaProGlyTrpSerArgValSerGlyIleProSerArgValHis 40
DB 109 CTTGGCTCGCGGTGCTCGGGGCTGCTCGGCTCGGGCTCGGCTCGGCTCGGCTCGG 168
QY 41 TrpProValProTrpYrtyrIysArgPheAspPheArgProIleProValMetGluIlyAspLeu 60
DB 169 TGCCCGGTGCTTCAACAGCCCTTGACTTCTGTCGCAAACTATCTTATTTGCAAGCT 228
QY 61 LysTrpThrPheCysProThrGlySerProIleProValMetGluIlyAspLeu 80
DB 229 AAGTATATCTTCTGTCTCACTGCTACCTATCCAGTTATGAGGGTATGATGACATT 288
QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheIysTrpGlyAspLeuGlyHis 100
DB 289 GAAGTTTTCGATTTACAGCCCGCAGTATGGGAATTTAAATATGAGAGACTTCCTGGGAC 348
QY 101 LeuIlyIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyIysAsnTrpThr 120
DB 349 TTGAATTTATGATGATGTCATTTGATTAAGATGATTAATTAATTAATTAATTAATTA 408
QY 121 MetGluTrpYrtyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisIleuArGProGlu 140
DB 409 ATGGAATGATGTAACCTTTCCAACTTGCAACTGATCAATTCCTCCGACTGAA 468

```

```

ZNRG:1111056
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Degenerate nucleotide sequence encoding zsls46
OTHER INFORMATION: polypeptide of SEQ ID NO:2
FEATURE:
NAME/KEY: variation
LOCATION: (1)...(1038)
OTHER INFORMATION: n is any nucleotide
C09-122-383-13

```

QY	I	MetArgaagGlyAlaGlyAlaAlaAlaArgGlyArgAlaSerTyrCysTrpAlaLeuAlaLeu	20
Db	1	ATGAGNMGNGAGNGCNGGNGCNGCNGNGNGNGMGNGCNSWTSGTGTGTGGCCNTTNGCNTYTN	60
QY	21	LeuTrpLeuAlaValAlaValProGlyTyrIspSerArgValSerGlyIleProSerArgGHis	40
Db	61	YTNNGYTNNGCNGTNGTNGTNCNGNGNTGGMNSMGNGTNGTNSGNAATHCCNMSNMGMNGNCAY	120
QY	41	TTrpProValProTyrTyrArgPheAspPheArgProGlyProAspProTyrCysGlnAla	60
Db	121	TGGCCNGTNCNTTAARMGNTTYCAATYTMGNCNNAACCGAATCCNAATATGTCARGCN	180
QY	61	LySTyrTrpPheCysProThrGlySerProIleProValMetGlyGlyAspAspAlle	80
Db	181	AARTRACNTTYYTGCCNAACNGMNSNCNAHCCNGTNAATGGARGCNGAYGAYATH	240
QY	81	GluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeuLeuGlyHis	100
Db	241	GARGTNTYMGNTYTCARGCNCNGTNTGGGARTTYAARTATAGMGAAATYTNYTNNGCAY	300
QY	101	LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr	120
Db	301	YTNARARATHAAGCAGAYGCNAATHGSGNTTYMGNSNACNTYTNACNGNAARARATYACN	360
QY	121	MetGluTrpTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGlu	140
Db	361	ATGGARGTGATGAYNTTYYCARLTNGNMAATYTGACNTTCCNCAATYTNMGNCNGAR	420
QY	141	MetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAsp	160
Db	421	ATGGAYGNCCTTYYTGGTGAYACARGGNGCNGCTGTATTTTTCARGGNAATHGAY	480
QY	161	ValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe	180
Db	481	GTNCAYTGGARGAARAAAYGCNACNTTNGNCARGTNGCNAACNAATHMSNGNMAATGTTY	540
QY	181	AsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyrTyrGluThrTrp	200
Db	541	AAYCARATGGCNAATGGGTNARCARGAYAYGARAACNGNATHATYATYAGRACNTGG	600
QY	201	AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSer	220
Db	601	AAYGTMARAGCWNSSCNGARARARGGNGGAPACNTGGTGTGATWSNTYATGATGWSN	660
QY	221	LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle	240
Db	661	AARTTYGNTYTMGNAACNTTAAYARATYTNNGCGARTTGGCNGCNGARTTYAARAAAYATH	720
QY	241	GluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrTrpLeuGlyAsnGlu	260
Db	721	GARACNAATYATACMNGNAHTTYYTNNTAYWSNGGARGCCNACNATYTNNGGNAAGAR	780
QY	261	ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr	280
Db	781	ACNWSGNTTYYGNCNCNACNGNANAYARACNTYTMGNTYTNCGCNAATHARMGNTTYAT	840
QY	281	TyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePhe	300
Db	841	TAYCCNTTAAACCMCAAYTNCNACNARGARATTYTNTYTNWSNTYTNACARATHTTY	900
QY	301	AspAlaValIleValHisLysGlnPheTyrLeuPheTyrAsnPheGluTyrTrpPheLeu	320



407 ATGGAAATGGTATGAACTTTTCCAACTTGGCAACTGTAACTTTCCCAATCTCCGACCTGAA 466

141 MetAapAlaProPheTIPCySaNgInGlyAlaAlaCySPhePheGluGlyIleAaSPaSP 160

467 ATGGATGCCCTTTCTCGGTGTATCAAGCCGCTGCTCTTTTGAAGGAATGATGAT 526

161 ValHISTrpLysGluSaNgIyThrLeuValGlnValAlaThrIleSerGlyAaMePhe 180

527 GTTCACTGGAGAAATGGGCACTTAATGTCAAGTGAACCTTAATACGGAACATGTTC 586

181 AaNgIaMeAlaIleTrpValLysGlnAaSPaNgIuThrGlyIleTyTyGluThrTrp 200

587 AACCAATATGCAGAGTGGGTGAACAGACCAATGAACAGAAATTTATTAAGACATGG 646

201 AaNuValLysAlaSerProGluLysGlyAlaGluThrTrpPheAaSPeSerTyraSPCySer 220

647 AATGTAAAGCCAGCCAGAAAGGGGCGAGACATGTTGATTCCTACGACGTGTTCC 706

221 LysPheValLeuArgThrPheAaSuLysLeuAlaGluPheGlyAlaGluPheLysAaMe 240

707 AAATTTGTGTAAAGCACTTTTAACAAGTTGGCTGAATTTGAGCGAGATTCAAGAACATA 766

241 GluThrAaSuTyThrArgIlePheLeuTySerGlyGluProThrTyLeuGlyAaSuGlu 260

767 GAACCAACTTAACAGAAATATTTCTTTACGTGAGAACTCACTATCTGGAAATGAA 826

261 ThrSerValPheGlyProThrGlyAaSuTyThrLeuGlyLeuAlaIleLysArgPheTy 280

827 AACTGTGTTTTGGGCAACAGAAACAAGACTCTGTGGTTTGGCATAAAGATTATAT 846

281 TyrProPheLysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePhe 300

847 TACCCCTTCAAAACCAATTTGCCAATAAAGAAATTTCTGTGAAGTCTCTGCAGAAATTTT 946

301 AspAlaValIleValHisLysGlnPheTyLeuPheTyraSPaNgIuTyTrpPheLeu 320

947 GATGCGATGATGTGACAAACAGTTCTATTTGTTTATTAATTTGAATATGGTTTTTA 1006

321 ProMeLysPheProPheIleLysIleThrTyGluGluIleProLeuProIleArgAa 340

1007 CCAATGAATTCCTCTTTATTAATAAATACATATGAAGAAATCCCTTACCTATCAGAAAC 1066

341 LysThrLeuSerGlyLeu 346

1067 AAAACACTCTCTGTGTTA 1084

RESULT 3

5-10-653-595-26

Sequence 26, Application US/10653595

Publication No. US20040048304A1

GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

FILE REFERENCE: P2027P1C1

CURRENT APPLICATION NUMBER: US/10/653,595

CURRENT FILING DATE: 2003-09-03

PRIOR APPLICATION NUMBER: US 09/397945

PRIOR FILING DATE: 1999-09-17

PRIOR APPLICATION NUMBER: PCT/US99/05804

PRIOR FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: 60/078,566

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,576

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,573

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,574

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/078,579

PRIOR FILING DATE: 1998-03-19

PRIOR APPLICATION NUMBER: 60/080,314

PRIOR FILING DATE: 1998-04-01

Seq	Accession	Length	Score	Alignment	Identity	Similarity	Gap
1	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
2	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
3	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
4	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
5	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
6	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
7	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
8	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
9	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
10	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
11	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
12	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
13	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
14	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
15	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
16	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
17	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
18	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
19	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
20	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
21	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
22	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
23	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
24	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
25	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
26	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
27	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
28	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
29	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
30	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
31	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
32	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
33	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
34	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
35	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
36	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
37	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
38	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
39	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
40	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
41	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
42	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
43	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
44	U01010	1751	3.73e-230	1751	99.42%	99.42%	0
45	U01010</						











QY	264	PhcGcIyPcRtGhNgIyAmLySfThrLeuGlyLeuAlaIlleLysArpHeTyrTrpProPhe	283	
Db	243	TTTGGGCCCAACGAGAAACAGACTCTTGTTGGTTTACGCAATTAATGATTTTATTCCTTC	302	
QY	284	LysPcHIsLeuPcRtHrIlySGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaVal	303	
Db	303	AAACACATTTGGCCACACTAAAGAAATTTCTGTGAGTCTCTTGCAAAATTTTGATGACGTG	362	
QY	304	IlleValHsLysGlnPheTyrLeuPheTyrAsnPhesLutyrTrpPheLeuPcRtMlys	323	
Db	363	ATTGTGCACAAACGTTCTATTTCTTTTAATAATTTAAATATTTGGTTTATTAACTAAGAA	422	
QY	324	PhePcRtHIsLeuPheLleThrTyrGluGlnIleProLeuProIleArgenLysfThrLeu	343	
Db	423	TTTCCTTTTATTAATAATTAACATATGAGAAATTCCTTTACTACTATGCAAAACAAACACTC	482	
QY	344	SerGlyLeu 346		
Db	483	TCTGTGTTTA 491		
RESULT 8				
	US-10-040-739-825			
	Sequence 825, Application US/10040739			
	Publication No. US20020173635A1			
	GENERAL INFORMATION:			
	APPLICANT: Jacobs, Kenneth			
	McCoy, John			
	LaVaille, Edward			
	Racie, Lisa			
	Merberg, David			
	Tracy, Maurice			
	Spaulding, Viki			
	TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS			
	NUMBER OF SEQUENCES: 1519			
	CORRESPONDENCE ADDRESS:			
	ADDRESSEE: Genetics Institute, Inc.			
	STREET: 87 CambridgePark Drive			
	City: Cambridge			
	STATE: Massachusetts			
	COUNTRY: U.S.A			
	ZIP: 02140			
	COMPUTER READABLE FORM:			
	MEDIUM TYPE: Floppy Disk			
	COMPUTER: IBM PC Compatible			
	OPERATING SYSTEM: PC-DOS/MS-DOS			
	SOFTWARE: PatentIn Release #1.0, Version #1.30			
	CURRENT APPLICATION DATA:			
	APPLICATION NUMBER: US/10/040,739			
	FILING DATE: 07-Jan-2002			
	CLASSIFICATION: <Unknown>			
	PRIOR APPLICATION DATA:			
	APPLICATION NUMBER: 09/036,520			
	FILING DATE: 03-JUN-1998			
	ATTORNEY/AGENT INFORMATION:			
	NAME: Brown, Scott A.			
	REGISTRATION NUMBER: 32,724			
	TELECOMMUNICATION INFORMATION:			
	TELEPHONE: (617) 498-8224			
	TELEFAX: (617) 876-5851			
	INFORMATION FOR SEQ ID NO: 825:			
	SEQUENCE CHARACTERISTICS:			
	LENGTH: 506 base pairs			
	TYPE: nucleic acid			
	STRANDEDNESS: double			
	TOPOLOGY: linear			
	MOLECULE TYPE: cDNA			
	SEQUENCE DESCRIPTION: SEQ ID NO: 825:			
	US-10-040-739-825			
Alignment Scores:				
	2.11e-81	Length:	506	
	Score:	727.00	Matches:	133

Percent Similarity:	100.00%	Conservative:	0
at Local Similarity:	100.00%	Mismatches:	0
ery Match:	37.73%	Indels:	0
:	14	Gaps:	0

-10-010-050A-2 (1-346) X US-10-040-739-825 (1-506)

151 AlaAlaCysPhePheGluGlyIleAspAlaValHisTyrLysGluAsnGlyThrLeuVal 170  
99 GCGCGCGCTTTTGGAGGAATGATGATGCTCACTGGAGGAAAATGGACATATGTT 158  
171 GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTyrValLysGlnAsp 190  
159 CAGACGACCACTAATATCAGGAACAATGTTCAACCAATGGCAAAAGTGATGAACAAGAC 218  
191 AsnGluThrGlyIleTyrTyrGluThrTyrAsnValLysAlaIleSerProGluLysGlyVala 210  
219 AATGAAACAGAAATTTATTTATGAGACATGGAAATGAAAGCCAGCCAGAAAAGGGGCA 276  
221 GluThrTyrPheAspSerTyrAspCysSerLysPheValLeuArgThrPheAsnLysLeu 230  
279 GAGCAATGGTTGATTCCTACGACTGTTCCAAATTTGTGTAAAGACCTTTAACAAATGG 338  
231 AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgLysPheLeuTyr 250  
339 GGTGAATTTGGAGCAGAGTTCAAGAACATGAAACCACTATACAGATATTTCTTAC 398  
251 SerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys 270  
399 AGTGAGAACCTAATCTATCTGGAAATGAAACATCTGTTTTTGGCCACAGAAACAG 450  
271 ThrLeuGlyLeuAlaIleLysArgPheThrTyrTyrProPhe 283  
459 ACTCTTGCTTTAGCCATAAAAAGATTTTATTAACCCCTTC 497

RESULT 9  
 S-09-864-761-5832  
 Sequence 5832, Application US/09864761  
 Patent No. US20020048763A1  
 GENERAL INFORMATION:  
 APPLICANT: Penn, Sharon G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wenheng  
 TITLE OR INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 FILE REFERENCE: Aeomic-a-X-1  
 CURRENT APPLICATION NUMBER: US/09/864,761  
 CURRENT FILING DATE: 2001-05-23  
 PRIOR APPLICATION NUMBER: US 60/180,312  
 PRIOR FILING DATE: 2000-02-04  
 PRIOR APPLICATION NUMBER: US 60/207,456  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: US 09/632,366  
 PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: GB 24263.6  
 PRIOR FILING DATE: 2000-10-04  
 PRIOR APPLICATION NUMBER: US 60/236,359  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: PCT/US01/00666  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00667  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00664  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00669  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00665  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00668  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00663  
 PRIOR FILING DATE: 2001-01-30

```

1  PRIOR APPLICATION NUMBER: PCT/US01/00662
2  PRIOR FILING DATE: 2001-01-30
3  PRIOR APPLICATION NUMBER: PCT/US01/00661
4  PRIOR FILING DATE: 2001-01-30
5  PRIOR APPLICATION NUMBER: PCT/US01/00670
6  PRIOR FILING DATE: 2001-01-30
7  PRIOR APPLICATION NUMBER: US 60/234,687
8  PRIOR FILING DATE: 2000-09-21
9  PRIOR APPLICATION NUMBER: US 09/608,408
10 PRIOR FILING DATE: 2000-06-30
11 PRIOR APPLICATION NUMBER: US 09/774,203
12 PRIOR FILING DATE: 2001-01-29
13 NUMBER OF SEQ ID NOS: 49117
14 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
15 SEQ ID NO 5812

```

Alignment Scores:			
Pred. No.:	4,24e-76	Length:	474
Score:	684.50	Matches:	129
Percent Similarity:	86.67%	Conservative:	1
Best Local Similarity:	86.00%	Mismatches:	1
Query Match:	35.52%	Indels:	19
DB:	9	Gaps:	1

US-10-010-050A-2 (1-346) x US-09-864-761-5832 (1-474)

US-10-010-050A-2 (1-346) x US-09-864-761-5832 (1-474)

Qy	153	CysphephegiuglyIlephaspValhIstirplysgIuasn glyThrleuValGlnVal	172
Db	82	TTTTTTTTTAA-----	93
Qy	173	AlaThrIleSerGIYAsmncPheasnGlnMetAlaYstirValIysgInaPasnGlu	192
Db	94	-----CTGGAAACATGTTTCAACCAAAAGGCAAAAGTGGGTGAACAGGCAATGAA	144
Qy	193	ThrgIyIleTyTrGluThrTrpAsnValIyGAlaSerProGluIyGegIyAlaGluThr	212
Db	145	ACAGGAATTTATTTATGAGACATGGAATGTAAAGCAGGCCGAAAGGGGGCAGAGACA	204
Qy	213	ThpPheaspSerTyAspCysSerIyPheValIleuAgtThrPheasnIyIleuAlaGlu	233
Db	205	TGGTTGATTCCTCCACATGTTCCAAATTTGGTTAAGACCTTTAAACAAGTGGCTGAA	264
Qy	233	PheGIyAlaGluPheIyAsnIleGluThrAsnTyThrArgIlePheIeuTySerGIY	253
Db	265	TTTGAGCGACAGTTCAAAGAACATAGAACCACTATACAAAGATTTCTTTACAGTGA	324
Qy	253	GluProThaTyIreugIyAsnGluThrSerValPheGIYProThrgIyAsnIyThIeu	277
Db	325	GAACCTACTTATCTGGAAATGAAACATCTGTTTGGGCCAACAGAAACAAGACTCTT	384
Qy	273	GIYleuAlaIleIyAsnArgPheTyTyProPheIyProIleuPocThrIyGluPhe	299
Db	385	GGTTTGCCCTAAAGATTTTATTAACCCCTTCAACACACTTGCCACTTAAGATTT	444
Qy	293	IeuIeuSerIeuIeuGlnIlePheAspAla	302
Db	445	CTGTTGAGTCTTCCAAATTTTGAATGA	474

10-264-049-1859  
 Sequence 1859, Application US/10264049  
 Publication No. US20040005579A1  
 GENERAL INFORMATION:  
 APPLICANT: Blise et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PA133P1  
 CURRENT APPLICATION NUMBER: US/10/264,049  
 CURRENT FILING DATE: 2002-10-04  
 PRIOR APPLICATION NUMBER: PCT/US01/18569  
 PRIOR FILING DATE: 2001-06-07  
 PRIOR APPLICATION NUMBER: US 60/209,467  
 PRIOR FILING DATE: 2000-06-07  
 NUMBER OF SEQ ID NOS: 4360  
 SOFTWARE: PatentIn Ver. 3.1  
 SEQ ID NO 1859  
 LENGTH: 497  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (2)..(2)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (175)..(175)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (285)..(285)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (295)..(295)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (313)..(313)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (323)..(324)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (331)..(331)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (395)..(395)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (440)..(440)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (458)..(458)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (464)..(464)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (478)..(478)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (483)..(483)  
 OTHER INFORMATION: n equals a,t,g, or c  
 3-10-264-049-1859

Alignment Scores:  
 Pred. No.: 2,51e-49 Length: 497  
 Score: 472.00 Matches: 88  
 Percent Similarity: 98.88% Conservative: 0  
 Best Local Similarity: 98.88% Mismatches: 1  
 Query Match: 24,49% Indels: 0  
 DB: 16 Gaps: 0

US-10-010-050a-2 (1-346) x US-10-264-049-1859 (1-497)

QY 258 GYAENGUJTHRSERVAJPHGIGYPROTHRGJYANLYBTHLEUGJLYEUAJALLEYS 277  
 DB 3 GGAATGAAACATCTGTTTGGCCCAAGGAAACAGACTCTGTTTACCCAAAA 62  
 QY 278 ARGPHETTYTYRPHOPHELYBPROHISLEUPROTHIRVSGJLPHLEUENSERLEU 297  
 DB 63 AGATTTTATMCCCTTCACCAACACTTGGCCACTTAAGAAATTTCTGTGAGTCTTG 122  
 QY 298 GJNLEPHEASPAJVALILEVALHISLYSGJNPHETTYRLEUPHETYZANPHGJUTYR 317  
 DB 123 CAATTTTGTGATGCGATGATTTGCAACAAAGTTCTATTGTTTATATNTGAAATY 182  
 QY 318 TYPHELEUPROMELYBEPHOPHEILELYSILETHTYTGJUGJULEPROLEUPRO 337  
 DB 183 TGGTTTTCCTATGAAATTCCTTTATTAAATAACATAGAGAAATCCCTTACT 242  
 QY 338 ILEARGANLYBTHLEUSERGJLYEU 346  
 DB 243 ATCAGAAACAAACACTCTGTGTTTA 269

# RESULT 11

US-09-867-701-2579  
 ; Sequence 2579, Application US/09867701  
 ; Patent No. US2002013237A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Aglate, Paul A.  
 ; APPLICANT: Jones, Robert  
 ; APPLICANT: Harlocker, Susan L.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
 ; FILE REFERENCE: 210121.497  
 ; CURRENT APPLICATION NUMBER: US/09/867,701  
 ; CURRENT FILING DATE: 2001-05-29  
 ; NUMBER OF SEQ. ID NOS: 10912  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2579  
 ; LENGTH: 351  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; US-09-867-701-2579

Alignment Scores:  
 Pred. No.: 2.06e-35 Length: 351  
 Score: 359.50 Matches: 78  
 Percent Similarity: 68.25% Conservative: 8  
 Best Local Similarity: 61.90% Mismatches: 25  
 Query Match: 18.66% Indels: 16  
 DB: 9 Gaps: 3

US-10-010-050a-2 (1-346) x US-09-867-701-2579 (1-351)

QY 112 SETHRLTHURHGYLYBAENTYRTHMEGLUTRPRYRGJULEUPHGLNLEUGJYASN 131  
 DB 3 AGTACATTAACTGGCAAGAACTTACACANTGAAATGRTAGTGAATTTCCAACTTGGCAAC 62  
 QY 132 CysThrRphProHISLEUARGPROGLUWELASPAJAPROPHETPCYASRGJGLYALA 151  
 DB 63 TGTACATTTCCCAATCTCCGACCTGGAATGATGATGCCCTTCTGTGTAAATCAAGGCCCT 122  
 QY 152 ALACysPhePheGJUGJYTLASAPSPYALHISTPULYSGJUAENGJYTHLEUVALGLN 171  
 DB 123 GC-TGCTTTTGAAGGGAATGATGATGATCTCACTGGAAGAAATGGACAATAGTCA 181



gment Scores: 0.00739 Length: 2667  
 re: 112.00 Matches: 84  
 cent Similarity: 32.17% Conservative: 36  
 t Local Similarity: 22.52% Mismatches: 121  
 ry Match: 5.81% Indels: 133  
 Gaps: 19

-10-010-050a-2 (1-346) x US-10-198-846-9753 (1-2667)

20 LeuLeuTriPheAlaValValProGlyTyrPheSer-----ArgValSer 33  
 1790 CTGCTCTGGCTCAGC-----CCAGCTTGAGTCTCCGCTGTCACTCTCCGCTGACCG 1737  
 34 GlyTleProSerArgArg-----HisTTPProValProTyrTyrAspPhePhe--- 50  
 1736 GCCTTCTCCCGACAGCGCTTCCACACAC-CCTGTGCTCTTGAGACCCAGCCCTGTGTA 1678  
 51 ArgProLysProAspProTyrCysGlnAlaLysTyrThrPhe----- 64  
 1677 AGTCCCTGGCCCGGCTCTCTGTGACACACACCGAGCTTCCAGAACACGAGACACC 1618  
 65 -----CysProThrGly----- 68  
 1617 GGGAGCTGACGTGGAGACAGACCTCTATGCGTGTCCCACTTCCCACTGAGTGAAT 1553  
 69 -----SerProIleProValMetGluGlyAsp-----AspAspIle 80  
 1557 GGGGACACATGACAGCCCGACAGCTTACAGCCATCTCTCACTGAGCTGGAGACAGTGC 1493  
 81 GluValPheArgLeuGlnAlaProValTyrGluPheLysTyrGlyAspLeuLeuGlyHis 100  
 1497 CAGCTATGTCTCTTCACTCCCG--TGGGATATG--TGGCAT 1459  
 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTyrThr 120  
 1458 CTCAGCCTTCTCCAGAGGTGGCATCTACAGTCCATGCTCCACAA-----ACG 1405  
 121 MetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGlu 140  
 1404 CAAAGGTGTCTCTTACTTAATA--AACGTCTCTTCCAGAACACAGCCCGCAA 1348  
 141 -----MetAspAlaProPhe----- 145  
 1347 CTCATCACTGTCAATAATGTGGCAAGGTGAAGCTCTGTGCTCCCTTACAGAGACAC 1288  
 146 ---TTPCysAsnGlnGlyAlaAlaCysPhePheGluGly----- 157  
 1287 GAGTGTGTCTGCAAGGACAGTGTCTGTGAGATGGAGATGGCCCTCAAGCTCCAA 1228  
 158 -----IleAspAspValHis--TTPGlyGluAsnGlyTyrLeuValGlnValAlaT 174  
 1227 ACCTGCCAAATACAGACATGTGACGGGCTCGGAGAGGGGTCTCTTGTCTCCATCCAGC 1168  
 174 IleIleSerGlyAsnMetCysPheAsnGlnMetAlaLys-----TTPValL 188  
 1167 GGGTCAAGCGGTCTCTTGGCGGGAGAAAGACCAACAGCCGCTTCCCTTCTG3--- 1122  
 188 yseGlnAspAsnGlnThrGlyTyrTyrGluThrTyrPheValLysAlaSerProGluL 208  
 1111 --TCACAGACAGACAGGCTTCAGAGAGAGCTGTGGGAAACATGTATGCCCCAT- 1055  
 208 yseGlyAlaGluThrTyrPheAspSerTyrAspCysSerLysPheValLysArgThrPhe 228  
 1054 --GGTAAACCTGGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1021  
 228 snLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIleP 243  
 1001 ----- 1021  
 248 heLeuTyrSerGlyLysProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrG 263

Db 1000 -----GCTG 997  
 Qy 268 LysnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuP 288  
 Db 996 GACCCCTGCTCTTGGGCTGACCGGCTCTTCTTCTTCTTCTTCAAGTTTCACTTCA 937  
 Qy 288 rothLysGluPhe-----LeuLeuSerLeuLeuGlnIlePheAspAlaValI 304  
 Db 936 TTGGCAGAAATTTCTTCTCTTAACTTCACTTGGCGCTTTCGCGCTTGGGCTTGC 877  
 Qy 304 LevalHisLysGlnPheTyrLeuPheTyrAsnPhe 315  
 Db 876 TCCTCTGACAGCAATTTATTTTTCATCTTTT 842  
 RESULT 15  
 US-10-369-493-33507  
 ; Sequence 33507, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xiandeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ. ID NOS: 47374  
 ; SEQ. ID NO 33507  
 ; LENGTH: 3530  
 ; TYPE: DNA  
 ; ORGANISM: *Desulfitobacterium hafnense*  
 US-10-369-493-33507  
 Alignment Scores:  
 Pred. No.: 0.0319 Length: 3530  
 Score: 108.50 Matches: 74  
 Percent Similarity: 38.80% Conservative: 42  
 Best Local Similarity: 24.75% Mismatches: 114  
 Query Match: 5.63% Indels: 71  
 Db: 16 Gaps: 15  
 US-10-010-050a-2 (1-346) x US-10-369-493-33507 (1-3530)  
 Qy 5 AlaGlyAlaAlaArgGlyArgAlaSerTyrCys-----TTPAlaLeuAla 19  
 Db 2224 GCATCGCGCGGCAAGAGGGGACCTATTATTATGACAGACAGCAATCCCTGGGAACTTCA 2283  
 Qy 20 LeuLeuTriPhe-----AlaValValProGlyTyrPheSerArgValSerGlyTyrPro 36  
 Db 2284 ATGGCTATATTTAGAGAAATGATCCATGGCCAGCGGGTGTCAAGTCTTCACTACATGAGG 2343  
 Qy 37 SerArgArgHisTTPProValProTyrTyrAspPheAspPheArgProLysProAspPro 56  
 Db 2344 AAAAAGCAAGAGCTCTTTAAACCGCAAAATGAAGACTTTGGCCCATGCGACCGCG 2403  
 Qy 57 TyrCysGlnAlaLysTyrThrPheCysProThrGlyLysSerProIleProValMetGluGly 76  
 Db 2404 CCAATTCCTATGTCATATCATCATGATCCCATGATGAGCAATTTTGGCTATCTCC----- 2457  
 Qy 77 AspAspAspIleGluValPheArgLeuGlnAlaProValTTPGluPheLysTyrGlyAsp 96  
 Db 2458 -----TTTATGTAGCGTTGGCCATTTGTGGAG-----GTCTCCTA- 2492  
 Qy 97 LeuLeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGly 116  
 Db 2493 GTGATTTGAGAGGTACCAATCTG-----ACCATCACCGCG 2528  
 Qy 117 LysAsn-----TyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsn 131

```
2529 ACCAATGCACTACACCTGGGAGACATTGCTTCTTTCTGCACTTGCCGAGTTTATC 2588
      |||
132 CysThrPheProHisLeuArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAla 151
      ::|
2589 ATGCCCATTTGGCAGAGTGCTCCACAGCTGTAATGTCATTATATGCTTTGGCAGGGGCG 2643
      ::|
152 AlaCysPhePheGlnGlyIleAspAspValHisTrpLysGluAsnGly-----ThrLeu 169
      |||
2649 GAACGGAATCTTCAACTCTGATGATAAAGCCGGAAGAGATGAGGCAAGTACCTTG 2703
      |||
170 ValGlnValAlaThrIleSerGly---AsnMetPheAsnGlnMetAla---LysTrpVal 187
      |||
2709 GTCCATGTCAA-ACAAGAAAGGGGCCCATATGTGAACCAAGCTCGCATACCGATGGGTG 2767
      |||
188 LysGlnAspAsnGluThrGlyIleTyrTyrGluThrTrpAsn-ValLysAlaSerProG 207
      ::|
2768 GGCTGGGAAATTCCTCAGCA-AGATGGCATGGTGGACTATCGGGAGCT 2815
      |||
207 uLysGlyAlaGluThrTrpPheAsp-----SerTyrAspCysSerLysPheValIle 224
      |||
2816 TAAAGGAGAAAGTCCGTTCTTGGATGTGACTTGGCTACGAGAGAAAGATGTTCT 2875
      |||
224 uArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTy 244
      |||
2876 GCAC----- 2879
      |||
244 rThrArgIlePheLeuTyrSerGlyGluProThrTyrIleuGlyAsnGluThrSerValPh 264
      |||
2880 ---AACATCACTTGTATGCC---GAACCC-----GGGCAAAAGGTGGCTTTTGT 2923
      |||
264 eGlyProThr-----GlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
      |||
2924 GGGGGCTACGGGAGCGGCAAGACCACTATTACCAATCTCAACCGTTTCTAT 2978
      |||
```

arch completed: April 25, 2004, 06:14:02  
b time : 697.597 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - nucleic search, using frame\_plus.p2n model  
on: April 24, 2004, 23:02:52 ; Search time 114,853 Seconds  
(without alignments)  
1671.819 Million cell updates/sec

le: US-10-010-050A-2  
1937  
fect score: 1 MRRGAGARGRASMCWALAL.....IKITYEIPRIRNKITSLG 346  
uence:

ring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

arched: 682709 seqs, 277475446 residues  
al number of hits satisfying chosen parameters: 1365418

imum DB seq length: 0  
imum DB seq length: 200000000  
it-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

mand line parameters:  
DEL-frame+ p2n.model -DEV=xlh  
/cgn2\_1/USPTO.spool/US10010050/runat.22042004.113205.27569/app.query.fasta\_...1372  
3-issued Patents NA -QFMT=fastap -SUFFIX=rm -TIMMARCH=0.1 -DOFCL=0  
3-issued Patents NA -QFMT=fastap -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=trans4.0.cdi  
IST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
JDB=LOCAL -OUTWMT=p2n -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
3ER=US10010050 @CGL 1.140 @runat.22042004.113205.27569 -NCPU=6 -ICPU=3  
J MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLBLOCK=100 -LONGIO  
J TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
3AEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

tebase :  
1: Issued Patents NA.\*  
2: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

ult No.	Score	Query Match	length	DB ID	Description
1	94.5	4.9	14070	4	US-09-108-006C-2
2	93	4.8	14643	4	US-09-453-702B-22
3	92	4.8	1827	4	US-09-489-039A-6332
4	91.5	4.7	7718	4	US-09-976-594-244
5	90.5	4.7	1349	1	US-08-290-448A-73
6	90.5	4.7	1349	1	US-08-290-448A-73
7	90.5	4.7	1349	1	US-08-175-069A-73
8	90.5	4.7	1349	4	US-08-461-939B-73
9	90.5	4.7	1349	4	US-08-461-939B-73
10	90.5	4.7	6418	1	US-08-464-000-73
11	90.5	4.7	6418	1	US-08-480-528A-11
12	90.5	4.7	6418	5	US-08-479-666-11
					PCT-US93-10520-11

13	90	4.7	3113	4	US-10-028-272-1
14	89.5	4.6	1328	1	US-08-290-448A-58
15	89.5	4.6	1328	1	US-08-290-448A-58
16	89.5	4.6	1328	1	US-08-175-069A-58
17	89.5	4.6	1328	4	US-08-461-939B-58
18	89.5	4.6	1328	4	US-08-464-000-58
19	89	4.6	12394	4	US-09-488-856A-10
20	88.5	4.6	4440	3	US-07-792-600-1
21	88.5	4.6	4440	3	US-09-157-021-1
22	88.5	4.6	4440	3	US-09-156-842-1
23	88.5	4.6	4440	3	US-09-591-514-1
24	88.5	4.6	5433	3	US-09-157-021-35
25	88.5	4.6	5433	3	US-09-156-842-35
26	88.5	4.6	5433	4	US-09-023-655-1363
27	88.5	4.6	5433	4	US-09-591-514-35
28	88.5	4.6	5433	4	US-09-620-312D-253
29	87.5	4.5	537	4	US-09-134-000C-219
30	86.5	4.5	13930	4	US-09-976-594-1011
31	86.5	4.5	13930	4	US-09-220-132-20
32	86	4.5	429	4	US-09-252-991A-6217
33	86	4.5	1398	4	US-09-252-991A-7174
34	86	4.5	1398	4	US-09-252-991A-6413
35	86	4.5	1575	4	US-09-252-991A-7330
36	86	4.5	1890	4	US-09-252-991A-6330
37	86	4.5	2166	3	US-09-232-197-101
38	86	4.5	2166	4	US-09-232-197-101
39	86	4.5	2166	4	US-09-232-201-101
40	86	4.5	2166	4	US-09-232-195-101
41	86	4.5	3577	4	US-09-620-312D-457
42	85.5	4.4	438	4	US-09-252-991A-9484
43	85.5	4.4	1533	4	US-09-252-991A-9378
44	85.5	4.4	2043	4	US-09-252-991A-9459
45	85.5	4.4	2274	4	US-09-252-991A-9347

ALIGNMENTS

RESULT 1  
US-09-108-006C-2  
Sequence 2, Appliication US/09108006C  
Patent No. 6524613  
GENERAL INFORMATION:  
APPLICANT: Steer, Clifford J.  
Kren, Belay T.  
Bandyopadhyay, Paramita  
Roy-Chowdhury, Tanya  
TITLE OF INVENTION: Hepatocellular Chimeraplasty  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kimeragen, Inc.  
STREET: 300 Pleasant Run  
CITY: Newtown  
STATE: PA  
COUNTRY: USA  
ZIP: 18940  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/108,006C  
FILING DATE: 30-Jun-1992  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/054,288  
FILING DATE: 30-APR-1997  
APPLICATION NUMBER: 60/054,837  
FILING DATE: 05-AUG-1997  
APPLICATION NUMBER: 60/064,996  
FILING DATE: 10-NOV-1997  
APPLICATION NUMBER: 60/074,497  
FILING DATE: 12-FEB-1998

```

QY      249 uUYrSerGylguPrrThrTYrLeuglYAsnGlUthSeRvAlpheglYPrroHnGLyAs 269
          |||:::|||||
Db      12549 TCAGGAAGCCACCAAGCAGTTTCCTCACGGAGACTCAAGGATAAcGTtTGATTGCCTTGGTAGC 12608
          |||:::|||||
QY      269 nlystrleuGlYleuAlAllelyS----- 277
          |||:::|||||
Db      12609 AGTTACTCAAGAATTCATCATATAAAGTCACAGACTCGATTGACTCATTGATTGATTTCT 12668
          |||:::|||||
QY      278 -----ArgPhetyrTYrProPhelysPrrHisleuPrrThrLYsGluPheleule 294
          |||:::|||||
Db      12669 GAACTTCCCCAGATTCACGTTTCGCCGGGAAAACCTGGGATATACACTAAGGAGAACTTTG 12728
          |||:::|||||
QY      294 uSerleuLeuGlnIllepheSpAlAlleVAlhSlYsGlnPheTYr----- 310
          |||:::|||||
Db      12729 CACTATGTTCC---ATTAAGGAGAGGTAGGAGACGGTACTGTCGCCAGGTATATTCAGAAAGTCCA 12785
          |||:::|||||
QY      311 -----leuPhetyrsnPheGlnuTYrTrpPheleuPrrMetLySPhePr 325
          |||:::|||||
Db      12786 TAATGTTCAGAAATACGTTTCTCTATTTCACAAGACCTAGATGATTACACTTCCTTCCA 12843
          |||:::|||||
QY      325 oPhelIlelyS 328
          |||
Db      12846 GTTAAGGAAA 12855

RESULT 2
US-09-453-702B-22
; Sequence 22, Application US/09453702B
; Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
           Burland, Valerie
           Perna, Nicole T.
           Plunkett, Guy
           Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSER: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: [Unknown]

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-Dec-1998

ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296,95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 4643
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-453-702B-22
Alignment Scores:
```



2y 207 ULys---GlyValGluThr-----ThrPheAspSerTyrAspSer----- 220  
 2b 1251 GAAATATGACCCCGACCTGTATGGCGTGTATGATGATATATATGCGCCGCGCTGTGCA 1310  
 2y 221 -LysPheValLeuArgThrPheAsnLeu-----AlaGluPhe----- 233  
 2b 1311 TCAGACAGAGCTGGCGAGCTCTTCAACAGACTGCACGCCATGATACGGCGCTTTTAATC 1370  
 2y 234 -----GlyAlaGluPheLysAsnLeuThrAsnTyrThrArgLeuPheLeuTyrSe 251  
 2b 1371 TCTTCTTGCAAAAGCTGAGCGCGCTGCGCTGCGTGTGTGCTGTTCAGCTGCGC 1430  
 2y 251 rGlyGluPro---ThrTyrLeuGlyAsnGluThrSerValPheGly-----Proth 267  
 2b 1431 GGGGGTACCGTGCATCTACTATGGCGAGAGTGGCGCTGATGCAATACATCCGTT 1490  
 2y 267 rGlyAsnTyrThrLeuGlyLeuAlaLeuArgPheTyrTyrProPheLysProHisLe 287  
 2b 1491 CTGCCGCAACCG-----TTCCCGTGGATCCGCGCT 1523  
 2y 287 uProThrLysGluPheLeuSerLeuGln 298  
 2b 1524 GCAGGATACCCAG---CTGCTGGCGCTGTATCAG 1554

RESULT 4  
 US-09-976-594-244  
 / Sequence 244, Application US/09976594  
 / Patent No. 6673549  
 / GENERAL INFORMATION:  
 / APPLICANT: Furness, Michael  
 / APPLICANT: Buchbinder, Jenny  
 / TITLE OF INVENTION: GENES EXPRESSED IN C3A LAYER CELL CULTURES TREATED WITH STEROIDS  
 / FILE REFERENCE: PA-0041 US  
 / CURRENT APPLICATION NUMBER: US/09/976,594  
 / PRIOR FILING DATE: 2001-10-12  
 / PRIOR APPLICATION NUMBER: 60/240,409  
 / NUMBER OF SEQ ID NOS: 1143  
 / SOFTWARE: PERL Program  
 / SEQ ID NO 244  
 / LENGTH: 7718  
 / TYPE: DNA  
 / ORGANISM: Homo sapiens  
 / FEATURE:  
 / NAME/KEY: misc\_feature  
 / OTHER INFORMATION: Incyte ID No. 6673549 322303.15  
 / NAME/KEY: unsure  
 / LOCATION: 6339-6362, 6967-7026  
 / OTHER INFORMATION: a, t, c, g, or other  
 US-09-976-594-244

Alignment Scores:  
 Pred. No.: 12.6 Length: 7718  
 Score: 91.50 Matches: 71  
 Percent Similarity: 31.40% Conservative: 37  
 Best Local Similarity: 20.64% Mismatches: 133  
 Query Match: 4.75% Indels: 103  
 DB: 4 Gaps: 15

US-10-010-050a-2 (1-346) x US-09-976-594-244 (1-7718)

QY 41 TrpProValProTyrTyrAsnArgPheAsp-----PheArgProLysProAspProTyrCys 58  
 DB 3443 TGGATGGCTCCGAACTATCTTGAACAAATCTACAGCACAGACGCGTGTGCT 3502  
 QY 59 GlnAlaLysTyr-----ThrPheCysProThrGlySerProIleProValMetGlu 75  
 DB 3503 TAGGAGATATGGCTGTGGAAATCTTCTCTTAGTGGGCTCTCATATACCCAGAGATACAA 3562  
 QY 76 GlyAspAspAspLe----- 60  
 DB 3563 ATGGATGAGGACTTTTGCAGTCCCTGAGGAAAGGATGAGATGAGAGCTCTGAGTAC 3622

QY 81 -----GlyValPheArgLeuGlnAlaProValTyrGlu----- 91  
 DB 3623 TCTACTCTCAAAATCTATCATGATCATGCTGAGCTGCAGACAGACCCAAAAGAG 3682  
 QY 92 -----PheLysTyrGlyAspLeuLeuGlyHisLeuLysIleMet 104  
 DB 3683 CCAAGATTTCAGAACTTGAGAAAACCTAGGTATGCTT---CAAGCAAAATGTACAA 3739  
 QY 105 HisAspAlaLeuGlyPheArgSerThrLeuThrGlyLysAsnTyrThrMetGluTyr 124  
 DB 3740 CAGAT-----GATTAAGCTACATCCCA-----ATC 3766  
 QY 125 GluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGluMetAspAlaPro 144  
 DB 3767 AATCCCATCTACAGAGAAATAGCGTTTACATCACTCACTGCTTCTTGAGAGAC 3826  
 QY 145 PheTyrCysAsnGlnGlyAlaAlaCysPhePheGlu-----GlyIleAspAspValHis 162  
 DB 3827 TTTCTCAAGAAAGTATATTCAGCTCCGAAAGTTTATTCAGAGAGCTTGATGATGTGCA 3886  
 QY 163 TryptylGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPheAsnGln 182  
 DB 3887 TATGTAAATGCTTCAAGTTTATGAGCTGGAAGAAATC-----AAACCTTGAGAA 3940  
 QY 183 MetAlaLysTyrValLysGlnAspAsnGluThrGlyIleTyr-----TyrGlu 198  
 DB 3941 CTTTTA-----CCGAATGCCACTCTCAATGTTGATGACTACACAGGCGAC 3985  
 QY 199 ThrTyrPheValLysAlaSerProGluLysGlyAlaGluThrTyrPheAspSerTyrAsp 218  
 DB 3986 AGCAGCACTGTGGCTCTCCATGCTGAGAGCGCTTCACCTGAGTACGAGCAAAACC 4045  
 QY 219 CysSer-----LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGly----- 234  
 DB 4046 AAGGCTTCGCTCAAGATTGACTTGAAGTACACAGTAAATAGAGATCGGGGCTGCT 4105  
 QY 235 -----AlaGluPheLysAsn 239  
 DB 4106 GATGTCAGAGGCCCAAGTTTCTGCATTCAGACTGTGGGCACTGCAAGGCAAGGCGC 4165  
 QY 240 IleGluThr----- 242  
 DB 4166 AGGTTCACTTACGACACAGCTGAGTGAAGAAATGCGGTGCTCCCGCCCA 4225  
 QY 243 AsnTyrThrArgIlePheLeuTyrSerGlyGluPro-----ThrTyr 256  
 DB 4226 GACTACAACTGGGTGCTCTGTACTCCACCCCATCTAGATTGACAGAGACCTT 4285  
 QY 257 LeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIle 276  
 DB 4286 ATTCTAGAGACACATGTGTAATTATACCCCGAGAAACTAGCTTTGCCAGATATTATGC 4345  
 QY 277 LysArgPheTyrTyrProPheLysProHisLeuProThrLysGluPheLeuSerLeu 296  
 DB 4346 ATATATTAAGTTTAAACCTTATCTTTCATGGAGACGACGCTTTTGTGATTTT 4405  
 QY 297 LeuGlnIlePhe 300  
 DB 4406 ATAGTCTTTT 4417

RESULT 5  
 US-08-290-448A-73  
 / Sequence 73, Application US/08290448A  
 / Patent No. 5676954  
 / GENERAL INFORMATION:  
 / APPLICANT: Rogers, Bruce  
 / APPLICANT: Klapper, David G.  
 / APPLICANT: Rafnar, Thorunn  
 / APPLICANT: Kuo, Mei-chang  
 / TITLE OF INVENTION: Allergenic Proteins From Ragweed and Usnea  
 / NUMBER OF SEQUENCES: 93  
 / CORRESPONDENCE ADDRESS:

```

ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IM1-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEO ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-290-448A-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93
DB: 1 Gaps: 16

US-10-010-050a-2 (1-346) X US-08-290-448A-73 (1-1349)
QY 65 CysPro-----ThGlySerProIleProValMetGluGlyAsp 77
DB 526 TGTCAGAGAGCGATGATTAGTCCAGCATGCTCCACCAATTTTAAAGACAAAGTGAT 585
QY 78 AspAspIleGluValPheArgGluGlnAlaProValTrpGluPheLysTyrGlyAspLeu 97
DB 586 GGTGATCTTAATGTTGCTGGTGAAGTTCACAAATATCG----- 624
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrIleuThrGlyLys 117
DB 625 ATGCACCATGCTGCTCAGTAAGGCTCCGATGGCTGCTCATATCACCCTCGGAGC 684
QY 118 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnGlySerThrPheProHisLeu 137
DB 685 TCACACGCTG-----ACCGTTTCCAACTGCCAAATTC----- 714
QY 138 ArgProIleuMetAspAlaProPheTrpCysAsnGlnGlyAlaIaCysPhePheGluGly 157
DB 715 -----ACCCACACCAATTTGATATTCCTGGG 744
QY 158 ILAspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
DB 745 GCGCATACACCCCTATTCAGAAATGAAGCATGCTA-----GCCAGGTAAGCATTC 795
QY 178 AsnMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyr--- 196

```

```

DB 796 AACATGTTCCACCATCAGCTGATGCCAAGAAATGCTAGATGATTTGGTTTTCGAA 855
QY 197 -----TyrGluThrTyrAsnValLysAlaSerProGluLysGlyAlaGlu 211
DB 856 GTCGTTAACAACTACGACAGATGGGAAAGTAAGCATGCTGATGCTGAGCCCA 915
QY 212 Thr-----TrpPheAspSerTyrAspCysSerLysPheValLeu 224
DB 916 ACTATACTCAGCCAAAGGAAAGACGATTCGCCCGCATGATATCATCAAG----- 966
QY 225 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 241
DB 967 -----AAAAATGCTTACGACGACTGTACTGGAACGACGAGATGCTGAGAAC 1020
QY 242 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThr 261
DB 1021 TCGAAGACAGATAGAGACTGCTGTGAATATGCT----- 1053
QY 262 SerValPheGlyProThrGly---AsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
DB 1054 GCTATTTTCTCCATCCGGGCTGATCCAGTCTAACCCCTGAGCAAAAGCAGGATG 1113
QY 281 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePhe 300
DB 1114 ATTCAGCTGACGACG-----GGAGAAAGCCGTTCTAGACTCACTAGTAGTGTCT 1161
QY 301 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 312
DB 1162 GGTGATCTCTCATGCGCATCAGAGACCTTGCTAGACCTGGCCAAATTCCTAGACTT 1221
QY 313 TyrAsn-----PheGluTyrTrpPheLeuProMetLysPheProHelleLysIle 329
DB 1222 TATATATCATTAATACTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1263
QY 330 ThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 343
DB 1264 TTAATATGA-----CATTAAGTTCAAGTACTCTA 1293

RESULT 6
US-08-290-448A-73
Sequence 73, Application US/08290448A
Patent No. 5698204
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207

```

```

REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-290-448A-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93
DB: 1 Gaps: 16

US-10-010-050A-2 (1-346) x US-08-290-448A-73 (1-1349)
QY 65 CysPro-----ThrglySerProIleProValMetGluGlyAsp 77
DB 526 TGTCAGAGAGGAGGATTAAGTCCAGCATGCTCCACCAATTTTAAGCAACAAAGTCAT 585
QY 78 AspAspIleGluValPheArgLeuGluAlaProValTrpGluPheLysTyrGlyAspLeu 97
DB 586 GGTATGCTATTAAGTGTGCTGATGCTACCAAAATATG----- 624
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
DB 625 ATCGACCATGTCCTCCCTCAGTAAGGCTTCGATGGGCTGTCGATATCAACCTCGGCGAGC 684
QY 118 AsnTrpThrMetGluTrpTyrGluLeuPheGluLeuGlyAsnCysThrPheProHisLeu 137
DB 665 TCACACGCG-----ACCGTTCCAACTGCATAATTC----- 714
QY 138 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
DB 715 -----ACCCAAACCAATTGTATTATGCTCGGG 744
QY 158 IleAspAspValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
DB 745 GCTGATGACACCCATTATCAAGATTAAGGCATGCTA-----GCAACGGTACATTC 795
QY 178 AsnMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyr--- 196
DB 796 AACATGTTCAACCGATCAAGTTCACCAAAAGATGCTAATGATGATGATTTGGTTTCCAA 855
QY 197 -----TyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGlu 211
DB 856 GTCGTTAACAACTACAGACAGATGGGGAACGTACGCATCGGTGATAGCTCGGCCCA 915
QY 212 Thr-----TrpPheAspSerTyrAspCysSerLysPheValLeu 224
DB 916 ACTATATCTCAGCCAAAGGAACAGATTCTTCCGCCCGCATGATATCATCAAG----- 966
QY 225 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 241
DB 967 -----AAAAATGCTTACCGAGAGACTGTACTGCGCAAGCAGAGATGCGTGGAAAC 1020
QY 242 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrIleuGlyAsnGluThr 261
DB 1021 TCGAAGAACGATAGAGACTTGTGAATGCT----- 1053
QY 262 SerValPheGlyProThrGly---AsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
DB 1054 GCTATTATTCTCCATCCGAGGTCTGATTCAGTGTCTAACCCCTGAGCAAAHAACAGGAGATG 1113

```

```

QY 281 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePhe 300
DB 1114 ATTCAGAGTGAACCA-----GGAGAAAGCCGTTCTAAGACTCACTAGTGTCT 1161
QY 301 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 312
DB 1162 GGTGTACTCTCATGATCATAGAGGACCTTGCTAAGACCTGGCCAAATCTTAGCTTT 1221
QY 313 TyrAsn-----PheGluTyrTrpPheLeuProMetLysPheProPheIleLysIle 329
DB 1222 TATTAATATCATTAATAACTTATTATTATTATT-----TTGGAAT 1263
QY 330 ThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 343
DB 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

RESULT 7
US-08-175-069A-73
Sequence 73, Application US/08175069A
Patent No. 5776761
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Ratnar, Thorum
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHYE & COCKFIELD, LLP
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/175, 069A
FILING DATE: December 29, 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529, 951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325, 365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-175-069A-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93

```

```

DB: 1 Gaps: 16
US-10-010-050a-2 (1-346) x US-08-175-069A-73 (1-1349)
QY 65 CysPro-----ThrglySerProIleProValMetGluGlyAsp 77
Db 526 TGTCCAGAGGAGCATGATTAAAGTCCAGATGGTCCCAATTAAAGACACAAAGTAT 585
QY 78 AspAspIleGluValPheArgLeuGlnAlaProValTrrGluPheLysTyrGlyAspLeu 97
Db 586 GGTGATGCTAAATGCTGCTGAGTTCACAAATATG----- 624
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
Db 625 ATGCACCATGCTCGCTCAGTAAGGCTCCGATGGCTCTCGATATTCACCTCGGCAGC 684
QY 118 AsnTyrThrMetGluTrrPyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 137
Db 685 TCACAGTG-----ACCGTTCCACTGCAATTC----- 714
QY 138 ArgProGluMetAspAlaProPheTrrCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
Db 715 -----ACCCACACCAATTGATATATGCTCGG 744
QY 158 IleAspAspValHisTrrPlyGluAbnGlyThrLeuValGlnValAlaThrIleSerGly 177
Db 745 GCTGATGACACCCCATTTATCAAGATAAGCATGCTA-----GCCACGGTAGCATTC 795
QY 178 AsnMetPheAsnGlnMetAlaLysTrrValLysGlnAspAsnGluThrGlyIleTyr--- 196
Db 796 AACATGTTCCAGCATTCAGTTGACCAAAGATGCCATGATGATGATGGGTTTCCAA 855
QY 197 -----TrrGluTrrPheAsnValLysAsnSerProGluLysGlyAlaGlu 911
Db 856 GTCGTTAACAACTACACAGATGGGGAACGATCCGCTGCTGAGTCTCGGCCCA 915
QY 212 Thr-----TrrPheAspSerTrrAspCysSerLysPheValLeu 224
Db 916 ACTATACTACGCCAAGGAAAGATGCTGCCCCGATGATATCATCAAG----- 966
QY 225 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 241
Db 967 -----AAAATGCTTAAAGGAGACTGTAAGTGGCAACGACAGTGCATGCTGAAC 1020
QY 242 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrIleGluLysAsnGluThr 261
Db 1021 TGGAGAACAGATAGAGACTGCTGAAATGCT----- 1053
QY 262 SerValPheGlyProThrGly---AsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
Db 1054 GCTATTTTCTCCATCCGGCTCGATCCAGTGTACCCCTGAGCAAAAAGCAGGATG 1113
QY 281 TyrThrPheLysProHisLeuProThrLysGlnPheLeuSerLeuLeuGlnIlePhe 300
Db 1114 ATTCAGCTGAACCA-----GGAGAACCCGTTCTTAAGACTCACAGTAGTGT 1161
QY 301 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 312
Db 1162 GGTGATGCTCAGTCAAGCAGACCTGCTAAGCACTGGCAGCAATTCCTAAGCTTT 1221
QY 313 TyrAsn-----PheGluTyrTrrPheLeuProMetLysPheProPheIleLysIle 329
Db 1222 TATATATATCATTAATATCTTATTTATTTATTT-----TTTGATATTT 1263
QY 330 ThrTrrGluGluIleProLeuProIleArgAsnLysThrLeu 343
Db 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTTA 1293

```

```

APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Ratner, Thorum
APPLICANT: Kuo, Mei-Chang
TITLE OF INVENTION: Methods For Treating Sensitivity To A
TITLE OF INVENTION: Protein Allergen Using Peptides which include A T Cell Epitope
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESS: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,939B
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/464,000
FILING DATE: 05-JUN-1995
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CNDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
FAX: (617)742-4214
INFORMATION FOR SEQ. ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-461-939B-73
Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conserved: 33
Best Local Similarity: 21.66% Mismatches: 120
Query Match: 4.70% Indels: 93
Gaps: 16
US-10-010-050a-2 (1-346) x US-08-461-939B-73 (1-1349)
QY 65 CysPro-----ThrglySerProIleProValMetGluGlyAsp 77
Db 526 TGTCCAGAGGAGCATGATTAAAGTCCAGATGGTCCCAATTAAAGACACAAAGTAT 585
QY 78 AspAspIleGluValPheArgLeuGlnAlaProValTrrGluPheLysTyrGlyAspLeu 97
Db 586 GGTGATGCTAAATGCTGCTGAGTTCACAAATATG----- 624
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
Db 625 ATGCACCATGCTCGCTCAGTAAGGCTCCGATGGCTCTCGATATTCACCTCGGCAGC 684
QY 118 AsnTyrThrMetGluTrrPyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 137

```

```

RESULT 8
US-08-461-939B-73
; Sequence 73, Application US/08461939B
; Patent No. 6335019
; GENERAL INFORMATION:

```



```

Db      685 TCACACGTG-----ACCGTTTCCAACTGCAGAAATTC----- 714
Qy      138 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
Db      715 -----ACCCACACCAATTTGATTATTTGCTCGGG 744
Qy      158 IleAspAspValHisTrpGlyGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
Db      745 GCTGATGACACCCATTATTCAGATTAAGGACATGCTA-----GCAACGGTAGCATTC 795
Qy      178 AsnMetPheAsnGlnMetAlaLysTrpValIleGlnAspAsnGluThrGlyIleTyr--- 96
Db      796 AACATGTTCCACCATTCACGTTGACCAAGATGCTGATGATGATTGGGTTTTCCAA 855
Qy      197 -----TyrGluThrTrpAsnValIleAlaSerProGluGlyGlyAlaGlu 211
Db      856 GTCGTTAACAAACACTACGACAGATGGGAAACGTACCGTGGTAGCTCGGCCCA 915
Qy      212 Thr-----TrpPheAspSerTyrAspCysSerLysPheValLeu 224
Db      916 ACTATCTACGACCCAGGAAACAGATTCTTGGCCCCCATGATATCATCAAG----- 366
Qy      225 ArgThrPheAsnLysLeuAlaGluPheGly-----AlaGluPheLysAsnIleGlu 341
Db      967 -----AAAAATGCTTAGGAGGAGCTGACTGCAACGAGAGCTGATGCTGGGAC 1020
Qy      242 ThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThr 261
Db      1021 TCGAGAACAGATAGAGACTTGTGAAATGCT----- 1053
Qy      262 SerValPheGlyProThrGly---AsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280
Db      1054 GCATTTTCTCCCATCGGGCTGTGATCCAGTCTTAACCCCTGAGCAAAAAGCAGGATG 1113
Qy      281 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuLeuGlnIlePhe 300
Db      1114 ATTCACAGCTGAACCA-----CGAGAGCGCTTCAAGACTCATAGTAGTGCT 1161
Qy      301 AspAlaValIleValHisLys-----GlnPheTyrLeuPhe 312
Db      1162 GGTGATCTCTGATGCATCAAGAGACCTTGTAGACCTGCGCATTTCTTAAGCTTT 1221
Qy      313 TyrAsn-----PheGluTyrTrpPheLeuProMetLysPheProPheIleLysIle 329
Db      1222 TATATTAATCATTAATCTATTATTATTATT-----TTTGATATTT 1263
Qy      330 ThrTyrGluGluIleProLeuProIleArgAsnLysThrLeu 343
Db      1264 TTATATGAA-----CCATTACGTTCAAGTACTCTTA 1293

RESULT 9
US-08-464-000-73
; Sequence 73, Application US/08464000
; Patent No. 6335020
; GENERAL INFORMATION:
; APPLICANT: Rogers, Bruce
; APPLICANT: Klapper, David G.
; APPLICANT: Rafnar, Thorunn
; APPLICANT: Kuo, Mei-chang
; TITLE OF INVENTION: Allergenic Peptides from Ragweed Pollen
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,000
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/290,448
FILING DATE: 15-AUG-1994
APPLICATION NUMBER: US 07/529,951
FILING DATE: 29-MAY-1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: 17-MAR-1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 1349 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1323
US-08-464-000-73

Alignment Scores:
Pred. No.: 0.96 Length: 1349
Score: 90.50 Matches: 68
Percent Similarity: 32.17% Conservative: 130
Best Local Similarity: 21.66% Mismatches: 93
Query Match: 4.70% Indels: 93
DB: 4 Gaps: 16

US-10-010-050a-2 (1-346) x US-08-464-000-73 (1-1349)
Qy      65 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 77
Db      526 TGTCCAGAGAGCATGATTAGTCCAAAGATGCTGCACAAATTTTAAGCAACAAAGATGAT 585
Qy      78 AspAspIleGluValPheArgLeuGlnAlaProValTrpGluPheLysTyrGlyAspLeu 97
Db      586 GGTATGCTTAATATGTTGCTGTAATTCACAAATATAG----- 624
Qy      98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
Db      625 ATGCACCATGCTGCTGCATAGAGCTTCCGATGAGCTGCTCCATATCACCCCTGGCAGC 684
Qy      118 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnGlySerThrPheProHisLeu 137
Db      685 TCACACGTG-----ACCGTTTCCAACTGCAGAAATTC----- 714
Qy      138 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
Db      715 -----ACCCACACCAATTTGATTATTTGCTCGGG 744
Qy      158 IleAspAspValHisTrpGlyGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
Db      745 GCTGATGACACCCATTATTCAGATTAAGGACATGCTA-----GCAACGGTAGCATTC 795
Qy      178 AsnMetPheAsnGlnMetAlaLysTrpValIleGlnAspAsnGluThrGlyIleTyr--- 196
Db      796 AACATGTTCCACCATTCACGTTGACCAAGATGCTGATGATGATTGGGTTTTCCAA 855
Qy      197 -----TyrGluThrTrpAsnValIleAlaSerProGluGlyGlyAlaGlu 211
Db      856 GTCGTTAACAAACACTACGACAGATGGGAAACGTACCGTGGTAGCTCGGCCCA 915
Qy      212 Thr-----TrpPheAspSerTyrAspCysSerLysPheValLeu 224

```

```

b 916 ACTATCTCAGCCAGGAAAGATCTCTGCCCCCGCATGATCATCAAG----- 966
y 225 ArgThrheanlysluallagluPhegly-----AlagluPheylsAsnileglu 241
b 967 -----AAAAATGCTTTCAGGAGACTGCTGATCTGCAACGAGAGTGCATCTGGAAC 1420
y 242 ThrAenlyrThrArgilePheLeuTyrserglygluProthryrLeuglyAsnluThr 261
b 1021 TGGAGAACAGATAGAGACTTCTTGAATAATGCT----- 1053
y 262 SerValPheglyProthrgly--AsnlyThrleuglyLeuAlailelyAsnlyrPheTyr 280
b 1054 GCTATTTTCTCCCATCCGGCTGATCCAGTGTACACCCCTGAGCAAAAAAGAGGATG 1.13
y 281 TyrProPheylsProHlsLeuProThryrleugluPheLeuSerLeuLeugluilePhe 300
b 1114 ATTCAGCTGACCA-----GGAAGAGCCGTTCTTAAGACTCAGTAAAGTGTCT 1.61
y 301 AspaAlaValileValHlslys-----GlnPheTyrleuPhe 312
b 1162 GGTGACTCTCATGCCATCAAGAGACCTTGCTAAGCAGCCTGGCCATTCCTAAGCTTT 1221
y 313 TyrAsn-----PhegluTyrTrpPheLeuProMetlyrPheProPheilelystle 329
b 1222 TATATATATCAATAATACTTATTATTATTATT-----TTGATATTT 1263
y 330 ThrTyrGluGluileProleuProileArgAsnlysThrleu 343
b 1264 TTATATGAA-----CCATTACGTTCAAGTACTCTA 1293

```

RESULT 10  
US-08-480-528A-11

; Sequence 11, Application US/08480528A  
; Patent No. 5652118

; GENERAL INFORMATION:  
; APPLICANT: OPPERMAN, HERMANN  
; APPLICANT: OKAYAK, ENGIN  
; APPLICANT: KUBERASAMPATH, THANAGAVEL  
; APPLICANT: RUBER, DAVID C.  
; APPLICANT: PANG, ROY H. L.  
; APPLICANT: COHEN, CHARLES M.  
; TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
; STREET: 45 SOUTH STREET  
; CITY: HOPKINTON  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 01748

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,528A  
; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FENTON Esq., GILLIAN M.  
; REGISTRATION NUMBER: 36,508  
; REFERENCE/DOCKET NUMBER: CRP-076FW  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 248-7560  
; TELEFAX: (617) 248-7100  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6418 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE: NAME/KEY: misc\_feature

LOCATION: 1..6361

OTHER INFORMATION: /note= "HOP-2 genomic sequence"

FEATURE: NAME/KEY: exon

LOCATION: 1..837

OTHER INFORMATION: /note= "EXON ONE"

FEATURE: NAME/KEY: misc\_feature

LOCATION: 884..885

OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN POSITIONS 884 AND 885 IN THIS SEQUENCE"

FEATURE: NAME/KEY: exon

LOCATION: 1088..1277

OTHER INFORMATION: /note= "EXON TWO"

FEATURE: NAME/KEY: exon

LOCATION: 1350..1814

OTHER INFORMATION: /note= "EXON THREE"

FEATURE: NAME/KEY: misc\_feature

LOCATION: 1834..1835

OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN POSITIONS 1834 AND 1835 IN THIS SEQUENCE"

FEATURE: NAME/KEY: exon

LOCATION: 1883..2077

OTHER INFORMATION: /note= "EXON FOUR"

FEATURE: NAME/KEY: exon

LOCATION: 2902..2981

OTHER INFORMATION: /note= "EXON FIVE"

FEATURE: NAME/KEY: exon

LOCATION: 3507..3617

OTHER INFORMATION: /note= "EXON SIX"

FEATURE: NAME/KEY: exon

LOCATION: 6116..6361

OTHER INFORMATION: /note= "EXON SEVEN"

US-08-480-528A-11

Alignment Scores:

Pred. No.: 12.3 Length: 6418

Score: 90.50 Matches: 67

Percent Similarity: 29.97% Conservative: 28

Best Local Similarity: 21.14% Mismatches: 97

Query Match: 4.70% Indels: 126

DB: 1 Gaps: 14

US-10-010-050A-2 (1-346) x US-08-480-528A-11 (1-6418)

```

QY 13 sETTPCySTrPAlaLeuAlaLeuLeuTrPLeuAlaValaProgly----- 28
b 532 TCTGTGGGCTGCGGCTATGCG---CGCTGGGGGGGGGGGGGGCGCGCGCGACCCCGC 588
y 29 -----TTPserArgValser---GlyleProserArgArgHls 40
b 589 CCGGCTGTCCCAAGCAGCTGTGGCGCGCGAGCGCGGAGCGCGAGCGGAGATCC 648
y 41 TTP-----ProValProTyrLysArgPheAspPheArgPro 52
b 649 TGGCGGTGCTCGGGCTGCTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
y 53 LysProAspProTyrCys-GlnAlaLysTyrThrPheCysProThnGlyserProilePr 72
b 694 CCGCGCGCTCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
y 72 C-ValMetGluGlyAspAspAsp----- 79

```

```

Db 748 AGCCATGCGCGGCGACGACGAGGAGGCGCGCCCGGAGCGCGCTTGGGCGGCG 807
Qy 79 ----- 79
Db 808 CCGACCTGTGATGAGCTTGTACATGAGTGAGTGCGGCGCGCGGAGCCCTGCG 667
Qy 80 --11leGlValPheArGleuGlnAlaProValTPrpGluPheArGlyTyrGlyAspLeuG 59
Db 868 GAGTAACTGCGCTGCGACGAGGCGCTTGTGGCTTACACCCCGGAGACCAAGCTGCG 527
Qy 99 1yHisLeuArGlyMetHisAspAlaTleGlyPheArGlySerThrLeuThrGlyLysAsnT 119
Db 928 AACAAAGTTTGACATAAATGAAGAGCGCGCCCGACCGAGCTTCTGGCTCCAC- 586
Qy 119 YrHrMetGluTPrpTyrGluLeuPheGlnLeuGlyAsnGlySerThrPheProHleuArG 139
Db 987 -----CTTGAAGTGGTGAGGCG- TGG-----GCGCGGTGGCTCAACACGCTGCGCC 1034
Qy 139 roGluMetAspAlaProPhe-TPrpCysAsnGlnGlyAlaAla-CysPhePheGlu----- 156
Db 1035 CTCCAGAGCCCGAGCCATTCTGAGTGGCAGCCGAGGCGCTGCTTGTCTTCTAGTGAGCG 1094
Qy 157 -----GlyTleAspAspValHisTPrpLysGlu----- 165
Db 1095 AGACCGTGCCTGGGCGACCGACCGCATTTGGAAGAGATTGCGCTTGAAGCTGACCGCA 1154
Qy 166 -----AsnGlyThrLeuValGlnValAla----- 173
Db 1155 GATCCCGGCTGGGAGGCGGCTGACAGCTGCGAGTTCCGATTTTACAAGTGCCAGCAT 1214
Qy 174 -----ThrIleSerGlyAsnMetPheAsnGlnMet----- 183
Db 1215 CCACCTGCTCAACAGAGACCTCCACGTCAGATGTTCCAGTGGTCCAGAGCAGTCCAA 1274
Qy 184 -----AlaLysTPrpVal 187
Db 1275 CAGTGCTTCCCTTGGCGCGGTCGCCACCTTAAACCCCGACCTCAAGTCTCATGTC 1334
Qy 187 1lyeGlnAspAsnGluThrGlyLeuTyrTyrGluThrTyrAsnValLysAlaSerProG 207
Db 1335 AAGGAGCCCGAGCG-----GAGTGGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1379
Qy 207 uLyGlyAlaGluThrTyrPheAspSerTyrAspCysSerLys 221
Db 1380 AGATGGGAAGATGCTTGGCCGAGGCGCTTCACTGTGGAG 1422

RESULT 11
US-08-479-666-11
; Sequence 11, Application US/08479666
; Patent No. 5652337
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OKKAINAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANAGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H. L.
; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,666

```

```

; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON Esq., GILLIAN M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: CRP-076DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 435-9001
; TELEFAX: (508) 435-6951
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6418 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..6361
; OTHER INFORMATION: /note= "nop-2 genomic sequence"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1..837
; OTHER INFORMATION: /note= "EXON ONE"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 884..885
; OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
; OTHER INFORMATION: POSITIONS 884 AND 885 IN THIS SEQUENCE"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1088..1277
; OTHER INFORMATION: /note= "EXON TWO"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1350..1814
; OTHER INFORMATION: /note= "EXON THREE"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1834..1835
; OTHER INFORMATION: /note= "A GAP OCCURS BETWEEN
; OTHER INFORMATION: POSITIONS 1834 AND 1835 IN THIS SEQUENCE"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1883..2077
; OTHER INFORMATION: /note= "EXON FOUR"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 2982..2981
; OTHER INFORMATION: /note= "EXON FIVE"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3507..3617
; OTHER INFORMATION: /note= "EXON SIX"
; FEATURE:
; NAME/KEY: exon
; LOCATION: 6116..6361
; OTHER INFORMATION: /note= "EXON SEVEN"
; US-08-479-666-11

Alignment Scores:
Pred. No.: 12.3 Length: 6418
Score: 90.50 Matches: 67
Percent Similarity: 29.97% Conservative: 28
Best Local Similarity: 21.14% Mismatches: 97
Query Match: 4.70% Indels: 126
DB: 1 Gaps: 14

US-10-010-050A-2 (1-346) x US-08-479-666-11 (1-6418)
Qy 13 SerTPrpCysTPrpAlaLeuAlaLeuLeuTPrpLeuAlaValAlaProGly----- 28
Db 532 TCTGTGGCTGGCGGCTATGCG---CGCTGGCGCGGCGGCGCGCGCGCTGCGAGCCCGCC 588

```

```

29 -----TTPSerArgValSer---GlyIleProSerArgArgHis 40
589 CCGGCTGTCCTCCAGCAGAGCTCTGGGCGCGCGCGCGCGAGCGAGCGAGATCC 648
41 TTP-----ProValProTyrIlyAspPheAspPheArgPro 52
649 TGGCGGTGCTGGGCTGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
53 LysProAspProTyrCys-GlnAlaIlyTyrPheCysProThrGlySerProIlePr 72
694 CCGCGCTGCTCCCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747
72 O-ValMetGluGlyAspAspAsp----- 79
748 ACGCCATGCGCGCGCGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 807
79 ----- 79
808 CCGAGCTGTCATGAGCTTCTTAAACATGGGTGAGTGGCGCGCGCGCGCGCG 867
80 --IleGluValPheArgLeuGlnAlaProValTyrGluPheIlyTyrGlyAspLeuG 93
868 GAGTAACTGGCTGAGCTGCGAGCGCGCTCTGCTGCTACACCGCGGAGCGAGCT 927
99 LysIleuIlySerIleMetHisAspAlaIleGlyPheArgSerThrIleuThrGlyIlyAsnT 119
928 AACAAACGTTGCACTAAATGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 936
119 YrThrMetGluTyrPyrGluIleuPheGlnIleuGlyAsnCysThrPheProHisleuA 139
987 -----CTTGAGTGGTGGGCTGGC-TGG-----GGCGGTGGCTCACACAGCTGTGCCCC 1334
139 roGluMetAspAlaProPhe-TripCysAsnGlnIlyAlaIle-CysPhePheGlu----- 156
1035 CTCACAGAGCCGAGCATTTCTGAGTGCAGCGCGCGCGCGCGCGCGCGCGCG 1094
157 -----GlyIleAspAspValHisTyrIlySerGlu----- 165
1095 AGACCGTGGCTGGCGCACACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1154
166 -----AsnGlyThrIleuValGlnValAla----- 173
1155 GATCCCGGCTGGAGGCGGCTACACAGCTGGGAGTTCCGGATTTCACAGGTGCCAGCAT 1214
174 -----ThrIleSerGlyAsnMetPheAsnGlnMet----- 183
1215 CCACCTGCTCAACAGAGACCTTCACGTCGACATGTTCCAGGTGTCACAGACAGTCCA 1274
184 ----- 187
1275 CAGGTGCTTCCCTTGGCGCGGTCGCCACCTAACCCCGCACCTTCACAGTCTCATGTGC 1334
187 llyAsnGlnAspAsnGlnThrGlyIleTyrTyrGluThrTrpAsnValIlyAlaSerProI 207
1335 AAGGCGAGCCGAGAG-----GAGTGTGTGGTGTAAAGAGAGCTCTCA 1379
207 ulysGlyAlaGluThrTrpPheAspSerTyrAspCysSerIlyS 221
1380 AGATGGGAAGATGCTTGGCCGAGGCGCTGCACCTGGGAAG 1422

```

```

COUNTRY: USA
ZIP: 01748
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10520
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/667,274
FILING DATE: 11-MAR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,764
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/753,059
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/752,857
FILING DATE: 30-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/923,780
FILING DATE: 31-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/922,813
FILING DATE: 31-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-076PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508)435-9001
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 6418 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..6361
OTHER INFORMATION: /note= "HOP-2 genomic sequence"
FEATURE:
NAME/KEY: exon
LOCATION: 1..837
OTHER INFORMATION: /note= "EXON ONE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 884..885
OTHER INFORMATION: /note= "A Gap Occurs Between
OTHER INFORMATION: Positions 884 and 885 in this Sequence"
FEATURE:
NAME/KEY: exon
LOCATION: 1088..1277
OTHER INFORMATION: /note= "EXON TWO"
FEATURE:
NAME/KEY: exon
LOCATION: 1350..1814
OTHER INFORMATION: /note= "EXON THREE"
FEATURE:
NAME/KEY: misc feature
LOCATION: 1834..1835
OTHER INFORMATION: /note= "A Gap Occurs Between
OTHER INFORMATION: Positions 1834 and 1835 in this Sequence"
FEATURE:
NAME/KEY: exon
LOCATION: 1883..2077
OTHER INFORMATION: /note= "EXON FOUR"
FEATURE:

```

## RESULT 12

PCT-US93-10520-11

Sequence 11, Application PC/TUS9310520

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: CREATIVE BIOMOLECULES, INC.

STREET: 45 SOUTH STREET

CITY: HOPKINTON

STATE: MA



```

APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A
FILING DATE: August 15, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/529,951
FILING DATE: May 29, 1990
APPLICATION NUMBER: US 07/325,365
FILING DATE: March 17, 1989
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-018CN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 1328 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1328
US-08-290-448A-58
Alignment Scores:
Pred. No.: 1.23 Length: 1328
Score: 89.50 Matches: 68
Percent Similarity: 29.45% Conservative: 28
Best Local Similarity: 20.86% Mismatches: 113
Query Match: 4.64% Indels: 117
DB: 1 Gaps: 15
US-10-010-050a-2 (1-346) x US-08-290-448A-58 (1-1328)
QY 65 CysPro-----ThGlySerProIleProValMetGluGlyAsp 77
DB 505 TGTCCAGAGGCGATTAAGTCCACAGATGTCACCAATTAAAGACAAAGAGGAT 364
QY 78 AsnAspIleGluValPheArgGluGlnAlaProValTrpGluPheLysTyrGlyAspLeu 37
DB 565 GGTATGCTTAATATGTTGCTGGTAGTTCACAAATATCG----- 503
QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117
DB 604 ATCCACCATTCGTCGTCAGTAAGGCTCCGATGGCTGCATATCACCCCTGGCAGC 663
QY 118 AsnTyrThrMetGluTyrTrpGluLeuPheGlnLeuGlyAsnGlyThrPheProHisLeu 137
DB 664 TCACACGCGT-----ACCGTTCCAACTGCMAATTC----- 693
QY 138 ArgProGluMetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGly 157
DB 694 -----ACCCAAACCAAAATTGATTAATGCTGGG 723

```

```

QY 158 IleAspAspValIleIleTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177
DB 724 GCTGATGACACCCATTATCAAGATTAAGGATGCTA-----GCAACGGTAGCATTC 774
QY 178 AsnMetPheAsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyr 196
DB 775 AACATGTTCCACCGATCAAGTTCACCAAGAAATGCTAGATGATTTGGGTTTTCGA 834
QY 197 -----TyrGluThrTrpAsnValLysAlaSerProGluLysGlyAlaGlu 211
DB 835 GTCTTAACAACACTACGACGATGGGAAACGTAACGCTACGTCGTAAGCTGGCCCCA 894
QY 212 Thr-----TrpPheAspSerTyrAspCysSerLys----- 221
DB 895 ACTATACACGACCAAGGACAGATTCCTGGCCCCGATGATATCATCAAGAAAAATGTC 954
QY 222 -----PheValLeuArgThrPheAsn 228
DB 955 TTAGCGAGACTGTACTGCGAACGACAGATCGATGCTGGAAGTGGAAACAGATAA 1014
QY 229 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 248
DB 1015 GACTTGCTTGAATAATGATGCT-----ATTTT 1041
QY 249 LeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerValPheGlyProThrGly 268
DB 1042 CTCACATCGGGCTGATCCAGTCAATCCCTGAGCAA----- 1080
QY 269 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuPro 288
DB 1081 -----AAAGCAGGAGATGCTCCAGCTCAACCA----- 1107
QY 289 ThrLysGluPheLeuLeuSerLeuGlnIlePheAspAlaValIleValHisLys 307
DB 1108 ---GGAGAGCCGTTTAAGACTCACTAGTACTGCTGCTGCTCATGCATCAAGGA 1164
QY 308 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 317
DB 1165 GCACCTTGCTAAGCAACCTGGCAATTCCTAAGCTTTATATATATCAATAACTATTT 1224
QY 318 TrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 337
DB 1225 TATTATTAT-----TTGATATTATTATATGAA-----CCA 1254
QY 338 IleArgAsnLysThrLeu 343
DB 1255 TTACGTTCAAGTACTCTA 1272
RESULT 15
US-08-290-448A-58
Sequence 58: Application US/08290448A
Patent No. 5698204
GENERAL INFORMATION:
APPLICANT: Rogers, Bruce
APPLICANT: Klapper, David G.
APPLICANT: Rafnar, Thorunn
APPLICANT: Kuo, Mei-chang
TITLE OF INVENTION: Allergenic Proteins From Ragweed and Uses
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:
ADDRESSER: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,448A

```

FILING DATE: August 15, 1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/529,951  
 FILING DATE: May 29, 1990  
 APPLICATION NUMBER: US 07/325,365  
 FILING DATE: March 17, 1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Amy E. Mandragouras  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: IMI-018CN  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617)227-7400  
 TELEFAX: (617)227-5941  
 INFORMATION FOR SEQ ID NO: 58:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1328 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1328  
 US-08-290-448A-58

Alignment Scores:  
 Pred. No: 1.23 Length: 1328  
 Score: 89.50 Matches: 68  
 Percent Similarity: 29.45% Conservative: 28  
 Best Local Similarity: 20.86% Mismatches: 113  
 Query Match: 4.64% Indels: 117  
 DB: 1 Gaps: 15

US-10-010-050A-2 (1-346) x US-08-290-448A-58 (1-1328)

QY 65 CysPro-----ThrGlySerProIleProValMetGluGlyAsp 77  
 DB 505 TGTCCAGGAGGCGATTAAGTCAACGATGCTCCACCAATTTTAAGCAACAAGATGAT 364  
 QY 78 AspAspIleGluValPheArgGluGluAlaIleProValTyrGluPheLysTyrGlyAspLeu 97  
 DB 565 GGTGATGCTATAAATGTTGCTGTAAGTTCACAAATATGG----- 603  
 QY 98 LeuGlyHisLeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLys 117  
 DB 604 ATCCGACATGTGCTGCTCAGTAAGGCTTCGATGGCTGCTGATATCAACCTCGGACG 663  
 QY 118 AsnTyrThrMetGluTyrTyrGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeu 137  
 DB 664 TCACACGGT-----ACCGTTCCAACTGCATAATTC----- 693  
 QY 138 ArgProGluMetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGly 157  
 DB 694 -----ACCCAAACCAATTTGATTAATGCTCGGG 723  
 QY 158 IleAspAspValHisTyrLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGly 177  
 DB 724 GCTGATGACACCCATTAATCAAGTAAGCATGCTA-----GCAACGGTACGATTC 774  
 QY 178 AsnMetPheAsnGlnMetAlaLysTyrValLysGlnAspAsnGluThrGlyTyrLeu 196  
 DB 775 AACATGTTACGACATCAAGTTCACCAAGATGCTAGATGATGATTTGGGTTTTCCAA 834  
 QY 197 -----TyrGluThrTyrPheAsnValLysAlaSerProGluLysGlyValGlu 211  
 DB 835 GTTCGTTAACACAACTACACAGATGGGACGTAACGTCATCGGTGTAAGTTCGGCCCA 894  
 QY 212 Thr-----TyrPheAspSerTyrAspCysSerLys----- 221  
 DB 895 ACTATATCTAGGCCAAGGACAGATTTCTGCCCCGATGATATCATCAAGGAAATATGTC 954  
 QY 222 -----PheValLeuArgThrPheAsn 228

DB 955 TTAGCGAGAGCTGCTACTGCGACAGCGAGTCAGTCTGGAACCTGGAGAACAGATAAA 1014  
 QY 229 LysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTyrThrArgIlePhe 248  
 DB 1015 GACTTGCTGAAATATGGTGT-----ATTTT 1041  
 QY 249 LeuTyrSerGlyGluProThrTyrThrLeuGlyAsnGluThrSerValPheGlyProThrGly 268  
 DB 1042 CTCCCATCGGGCTCGATCCAGTGTCAACCCCTGAGCAA----- 1080  
 QY 269 AsnLysThrLeuGlyLeuAlaIleLysArgPheTyrTyrProPheLysProHisLeuPro 288  
 DB 1081 -----AAAGCAGGATGATTCACGCTGAACCA----- 1107  
 QY 289 ThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaValIleValHisLys 307  
 DB 1108 ---GGAGAAAGCGGTCTTAAGACTCACTAGTAGTGTGGTACTCTCATGCGCATCAAGGA 1164  
 QY 308 -----GlnPheTyrLeuPheTyrAsn-----PheGluTyr 317  
 DB 1165 GCACCTTGCTAAGCAACCTGCGCAATTCCTAAGCTTTATATAATCATATAATCTTATT 1224  
 QY 318 TrpPheLeuProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuPro 337  
 DB 1225 TATTTTATT-----TTTGATATTATATGAA-----CCA 1254

Search completed: April 25, 2004, 05:50:25  
 Job time: 148.853 secs





XX Sheppard PO, Gilbertson DG;  
 PI  
 XX  
 DR MPI; 1999-142930/12.  
 P-PSDB; AAW92967.  
 XX  
 PT New secreted polypeptide, zsig46, and its fragments, related fusion  
 PT proteins - used for diagnosis and treatment of thyroid disorders or  
 PT diseases involving genes on chromosome 13.  
 XX  
 PS Claim 27, Page 88-90; 101pp; English.

CC This invention describes the isolation of a novel human secreted protein,  
 CC zsig46 encoded by a gene on chromosome 13 which is mainly expressed in  
 CC the thyroid. This product can be used to study secretion of proteins from  
 CC cells and also to treat or prevent deficient expression of zsig46, which  
 CC may be associated with thyroid diseases (e.g. hypothyroidism, Graves'  
 CC disease, thyrotoxicosis, thyroid cancer etc.) or with diseases that  
 CC involve genes in the same region of chromosome 13 (e.g. Hirschsprung's  
 CC disease, neuronal ceroid-lipofusiosis, Wilson disease and Reiger  
 CC syndrome). Antibodies and other binding proteins, are used as immunosay  
 CC reagents to detect zsig46 or cells expressing it, e.g. for assessing  
 CC thyroid function to produce anti-idiotypic antibodies, for affinity  
 CC purification of zsig46, to screen expression libraries, to neutralise  
 CC zsig46 activity, and to deliver toxins, radioisotopes etc. for  
 CC therapeutic or diagnostic purposes. Agonists of the product can be used  
 CC to promote growth, differentiation and proliferation of specific cell  
 CC types, e.g. for treating (extra)thyroid diseases or as additive to cell  
 CC cultures

SO Sequence 1486 BP; 432 A; 302 C; 325 G; 427 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 4.89e-194 Length: 1486  
 Score: 1927.00 Matches: 346  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 2

US-10-010-050A-2 (1-346) X AAX02855 (1-1486)

QY 1 MetProValProTyrTyrValArgPheAspPheArgProValProTyrTyrPheValLeu 20  
 DB 47 ATGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 106  
 QY 21 LeuTrpLeuAlaValAlaProGlyTyrSerArgValSerGlyIleProSerArgArgHis 40  
 DB 107 CTTTGGCTCGGGGCTTCGGGGCTTCGGGGCTTCGGGGCTTCGGGGCTTCGGGGCTTCGG 166  
 QY 41 TrpProValProTyrTyrValArgPheAspPheArgProValProTyrTyrPheValLeu 60  
 DB 167 TGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 226  
 QY 61 LysTyrTrpPheCysProThiGlySerProIleProValMetGluGlyAspAspAspIle 80  
 DB 227 AAGTAACTCTTGTCTCAACTGCTCACTTCCAGTTAGAGGGGAGTGAATGACATT 286  
 QY 81 GluValPheArgLeuGlnAlaProValTyrGluPheTyrGlyAspLeuLeuGlyHis 100  
 DB 287 GAAAGTTTTCATTCAGAGCCCAAGCCCAATTAATTAATTAATTAATTAATTAATTAAT 346  
 QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerTrpLeuThiGlyLysAsnTyrThr 120  
 DB 347 TTGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 406  
 QY 121 MetGluTyrPyrGluLeuPheGlnLeuGlyAsnGlySerTrpPheProHisLeuArgProGlu 140  
 DB 407 ATGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 466  
 QY 141 MetAspAlaProPheTyrCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAsp 160  
 DB 467 ATGGATGCGCTTTTGGTGTATCAAGCGCTGCTGCTTTTGGAGGAATGATGAT 526

QY 161 ValHisTrpLysGluAsnGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetPhe 180  
 DB 527 GTTCACTGGAAGGAAATGGACATTAAGTCAAGTAGCACTATATCAGAAACATGTTTC 586  
 QY 181 AsnGlnMetAlaLysTrpValLysGlnAspAsnGluThrGlyIleTyrTyrGluThrTrp 200  
 DB 587 AACCAATGGCAAGTGGTGAACACGACATGAAACAGAAATTAATTAATTAATTAATTA 646  
 QY 201 AsnValLysAlaSerProGluLysGlyAlaGluThrTrpPheAspSerTyrAspCysSer 220  
 DB 647 AATGTAAAGCCAGCCCAAGAAAGGGGCGAGACATGTTTGAATCTCAAGATGTTCC 706  
 QY 221 LysPheValLeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIle 240  
 DB 707 AATTTGTTTGAAGACCTTTTAACTGCTGATTTTGGACAGAGTTCAAGAACATTA 766  
 QY 241 GluThrAsnTyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGlu 260  
 DB 767 GAAACCACTATACAGAAATATTTCTTACAGTGGAGAACTTACTTATCTGGGAAATGA 826  
 QY 261 ThrSerValPheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysArgPheTyr 280  
 DB 827 AACTCTTTTGGGCGCAACAGAAACAGACTCTTGTGTTAGCCATAAAGATTTTAT 886  
 QY 281 TyrProPheLysProHisLeuProThrLysGluPheLeuSerLeuGlnIlePhe 300  
 DB 887 TACCCCTTCAACCACTATGGCAACTTAAAGAAATTTCTTGAATCTTGGCAAAATTTT 946  
 QY 301 AspAlaValIleValHisLysGlnPheTyrLeuPheTyrAspPheGluTyrTrpPheLeu 320  
 DB 947 GATGACGATATTTGACAAACAGTCTTATTTGTTTAAATTTTGAATTTGATTTTAA 1006  
 QY 321 ProMetLysPheProPheIleLysIleThrTyrGluGluIleProLeuProIleArgAsn 340  
 DB 1007 CCTATAAATTCCTTTTATTAATAACATTAAGAAATCCTTACTATCAGAAAC 1066  
 QY 341 LysThrLeuSerGlyLeu 346  
 DB 1067 AAAACACTCTCTGTTTA 1084

RESULT 2  
 AA224826  
 ID AA224826 standard; DNA; 1751 BP.  
 XX  
 AC AA224826;  
 XX  
 DT 02-DEC-1999 (first entry)  
 XX  
 DE Human secreted protein gene 16 clone HMZAD77.  
 XX  
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;  
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;  
 KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;  
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W09947540-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 EF 18-MAR-1999; 59WO-US005804.  
 XX  
 PR 19-MAR-1998; 98US-0078563P.  
 PR 19-MAR-1998; 98US-0078566P.  
 PR 19-MAR-1998; 98US-0078573P.  
 PR 19-MAR-1998; 98US-0078574P.  
 PR 19-MAR-1998; 98US-0078576P.







CC Libraries that were synthesised by the oligo-capping method. The primers  
 CC enable the production of the full length cDNA easily without any special  
 CC methods. The present sequence was used as the representative sequence  
 CC from a human clone which was used in homology searches to identify the  
 CC clone. Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in CD-ROM format directly from  
 CC EPO

XX Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;

## Alignment Scores:

Pred. No.:	Length:	Matches:
4.22e-103	697	200
Score:	1065.00	Conservative: 3
Percent Similarity:	93.55%	Mismatches: 8
Best Local Similarity:	92.17%	Indels: 6
Query Match:	55.27%	Gaps: 2

DB: 4

US-10-010-050A-2 (1-346) x AAK93424 (1-697)

QY 1 MetArgArgGlyAlaGlyAlaAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20  
 DB 52 ATGCGGCGGGGCGCGCGCGCTCGGGGACCGCGCTTCTGCTGCTGGGCGCTCGCGCTG 111  
 QY 21 LeuTrpLeuAlaValValProGlyTrpSerArgValSerGlyIleProSerArgArgHis 40  
 DB 112 CTTTGGCTCGCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCT 171  
 QY 41 TrpProValProTrpTrpArgPheArgPheArgProGlyProGlyProGlyProGlyProGly 60  
 DB 172 TGGCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCT 231  
 QY 61 LysTrpTrpPheCysProTrpGlySerProGlyProGlyProGlyProGlyProGlyProGly 60  
 DB 232 AAGTATACCTTCTGCTCAACTGCTCACTGCTCACTGCTCACTGCTCACTGCTCACTGCT 291  
 QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheArgTrpGlyLysPheLeuGlyHis 100  
 DB 292 GAAGTTTTCGATTACAGCCCAAGTATGGAATTTAATATGGAACCTCCCGGACAC 351  
 QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTrpThr 120  
 DB 352 TTGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411  
 QY 121 MetGluTrpTrpGluLeuPheGlnLeuGlyAsnCysThrPheProHisLeuArgProGlu 140  
 DB 412 ATGGAAATGATGAACTTTTCCAACTTGCACTGATCATTTCCCATCTCCGACCTGAA 471  
 QY 141 MetAspAlaProPheTrpCysAsnGlnGlyAlaAlaCysPhePheGluGlyIleAspAsp 60  
 DB 472 ATGGATCCCTCTTCTGCTGATTCAGAGCGCTGCTGCTTTCCTTTCAGGGAATTGATGAT 531  
 QY 161 ValHisTrpLysGluLeuGlyThrLeuValGlnValAlaThrIleSerGlyAsnMetP 80  
 DB 532 GTTCACAGGAAGAAATGGGACATTAAGTTCAAGTCAACATATATCAAGAAACATGT 591  
 QY 180 heaengimetalal-lystrpvallygln---aspsenglunthrglyllelytrpGlu 198  
 DB 592 TCACACCAATGGCAAAAGTGGTGAAACAGAGCAATNGAAACAGGAATTTATATGAG 551  
 QY 199 -ThTrpAsnValLysAlaSerProGluLys---GlyAlaGlu 211  
 DB 652 AACATGGAATTTTAAAGCCANCCANBAAABAAAGGGGCAAAA 694  
 RESULT 6  
 ID AAK92158 standard; cDNA; 697 BP.  
 XX AAK92158;  
 AC AAK92158;  
 XX 06-NOV-2001 (first entry)  
 DT 06-NOV-2001 (first entry)  
 XX Human cDNA 5'-end sequence, SEQ ID NO: 618.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-00114089.

XX 08-JUL-1999; 99JP-00194486.

XX 11-JAN-2000; 2000JP-00118774.

XX 02-MAY-2000; 2000JP-00183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX WPI: 2001-524255/58.

PT 830 Primers useful for synthesizing full length cDNA clones and their use  
 in genetic manipulation.

Claim 2; SEQ ID NO 618; 1380bp + Sequence listing; English.

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the full length cDNA are useful for determining the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is the nucleotide sequence of the 5'-end of a cDNA provided in the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

XX Sequence 697 BP; 183 A; 160 C; 186 G; 163 T; 0 U; 5 Other;

## Alignment Scores:

Pred. No.:	Length:	Matches:
4.22e-103	697	200
Score:	1065.00	Conservative: 3
Percent Similarity:	93.55%	Mismatches: 8
Best Local Similarity:	92.17%	Indels: 6
Query Match:	55.27%	Gaps: 2

DB: 4

US-10-010-050A-2 (1-346) x AAK92158 (1-697)

QY 1 MetArgArgGlyAlaGlyAlaAlaAlaArgGlyArgAlaSerTrpCysTrpAlaLeuAlaLeu 20  
 DB 52 ATGCGGCGGGGCGCGCGCGCTCGGGGACCGCGCTTCTGCTGCTGGGCGCTCGCGCTG 111  
 QY 21 LeuTrpLeuAlaValValProGlyTrpSerArgValSerGlyIleProSerArgArgHis 40  
 DB 112 CTTTGGCTCGCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCT 171  
 QY 41 TrpProValProTrpTrpArgPheArgPheArgProGlyProGlyProGlyProGlyProGly 60  
 DB 172 TGGCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCTCGGCTGCT 231  
 QY 61 LysTrpTrpPheCysProTrpGlySerProGlyProGlyProGlyProGlyProGlyProGly 80  
 DB 232 AAGTATACCTTCTGCTCAACTGCTCACTGCTCACTGCTCACTGCTCACTGCTCACTGCT 291  
 QY 81 GluValPheArgLeuGlnAlaProValTrpGluPheArgTrpGlyLysPheLeuGlyHis 100  
 DB 292 GAAGTTTTCGATTACAGCCCAAGTATGGAATTTAATATGGAACCTCCCGGACAC 351  
 QY 101 LeuLysIleMetHisAspAlaIleGlyPheArgSerThrLeuThrGlyLysAsnTrpThr 120





PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PI  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX PT gene expression in human fetal liver.  
XX  
XX Claim 4; SEQ ID NO 19091; 639pp + Sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for measuring  
XX CC human gene expression in a sample derived from human foetal liver. The  
XX CC single exon nucleic acid probes may be used for predicting, measuring and  
XX CC displaying gene expression in samples derived from human fetal liver. The  
XX CC present sequence is a single exon nucleic acid probe of the invention.  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

#### Alignment Scores:

Pred. No.:	2,45e-84	Length:	494
Score:	886.00	Matches:	163
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	45.98%	Indels:	0
DB:	4	Gaps:	0

US-10-010-050A-2 (1-346) x ABA70786 (1-494)

```
QY 184 AAlaySTRPVallyLysGlnASPasnGluThrGlylleTyTYrGluThrTPAsnVallys 203
DB 3 GCAAAAGTGGGTGAACAGACGACATGAAACAGAAATTTATGTGACATGAAATGTAAA 62
QY 204 AAlaserProGluYsglyAlaGluThrTPheAspSerTYrAspCYseSerlyspheVal 223
DB 63 GCCAGCCCAAGAAAGGGGGGAGAGACATGTTGATTCCTACGACTGTTCCAAATTTGTG 122
QY 224 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLysGluThrAsn 243
DB 123 TTAAGGACCTTTAAACAAGTTGGCTGAATTTGGAGAGAGTTCAAGAACTAGAAACCAAC 182
QY 244 TYrThrArgGilePheLeuTYrSerGlyGluProThrTYrLeuGlyAsnGluThrSerVal 263
DB 183 TATACAGAAATATTTCTTTACAGTGGAGACCTTATCTGCGAAATGAAACATCTGTT 242
QY 264 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysAsnGlyPheTYrProhe 283
DB 243 TTTGGGCCAAGAGAAACAGAACTCTGGTTAGCCATTAAGATTTATATACCCCTTC 302
QY 284 LysProHisLeuProThrLysGluPheLeuLysSerLeuLeuGluIlePheAspAlaVal 303
DB 303 AAACCACTTTGCCAATCAAGAAATTTCTGTTGAGCTCTTCCAAATTTTATGCAAGTG 362
QY 304 IleValHisLysGluPheTYrLeuPheTYrAsnPheGluTYrTrpPheLeuProheLys 323
DB 363 ATTTGGCAACAACAGTTCTATTTGTTTATATTTGAAATTTGGTTTATCTATGAAA 422
QY 324 PheProPheIleLysIleThrTYrGluGluIleProLeuProIleArgAsnLysThrLeu 343
DB 423 TTCCCTTTATTTAAATTAACATATGAAATCCCTTTACTATCATGAAACAAACACATC 482
QY 344 SerGlyLeu 346
DB 483 TCTGTTTAA 491
```

RESULT 9

```
AA150961
ID AA150961 standard; DNA; 494 BP.
XX
XX AA150961;
XX AC
XX 17-OCT-2001 (first entry)
XX DT
XX DE Probe #19647 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD
XX 09-AUG-2001.
XX PF
XX 30-JAN-2001; 2001WO-US000663.
XX PR
XX 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI
XX WPI; 2001-488937/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX
XX Claim 25; SEQ ID NO 19647; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders
XX
XX SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;

Alignment Scores:



|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 2,45e-84 | Length:       | 494 |
| Score:                 | 886.00   | Matches:      | 163 |
| Percent Similarity:    | 100.00%  | Conservative: | 0   |
| Best Local Similarity: | 100.00%  | Mismatches:   | 0   |
| Query Match:           | 45.98%   | Indels:       | 0   |
| DB:                    | 4        | Gaps:         | 0   |



US-10-010-050A-2 (1-346) x AA150961 (1-494)



```
QY 184 AAlaySTRPVallyLysGlnASPasnGluThrGlylleTyTYrGluThrTPAsnVallys 203
DB 3 GCAAAAGTGGGTGAACAGACGACATGAAACAGAAATTTATGTGACATGAAATGTAAA 62
QY 204 AAlaserProGluYsglyAlaGluThrTPheAspSerTYrAspCYseSerlyspheVal 223
DB 63 GCCAGCCCAAGAAAGGGGGGAGAGACATGTTGATTCCTACGACTGTTCCAAATTTGTG 122
QY 224 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnLysGluThrAsn 243
DB 123 TTAAGGACCTTTAAACAAGTTGGCTGAATTTGGAGAGAGTTCAAGAACTAGAAACCAAC 182
QY 244 TYrThrArgGilePheLeuTYrSerGlyGluProThrTYrLeuGlyAsnGluThrSerVal 263
DB 183 TATACAGAAATATTTCTTTACAGTGGAGACCTTATCTGCGAAATGAAACATCTGTT 242
QY 244 TYrThrArgGilePheLeuTYrSerGlyGluProThrTYrLeuGlyAsnGluThrSerVal 263
DB 183 TATACAGAAATATTTCTTTACAGTGGAGACCTTATCTGCGAAATGAAACATCTGTT 242
```


```

Y 264 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLeuAspPheTyrTyrProPhe 283  
 b 243 TTTGGGCGCAACAGAAACAAAGACTCTGTGTTAGCCATAAAAGATTTATTAACCCCTTC 302  
 Y 284 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaVal 303  
 b 303 AAACCAATTGGCCAACTAAAGAAATTTCTGTGAGTCTCTTGCAATTTTGATGCAGTG 362  
 Y 304 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGlnTyrTyrPheLeuProMetLys 323  
 b 363 ATTGTGCACAAACAGTCTCTATTGTTTATATATTTGAATTTGTTTAACTATGAAAA 422  
 Y 324 PheProPheIleLysIleThrTyrGluGlnIleProLeuProIleAspAsnLysThrLeu 343  
 b 423 TTCCCTTTTATTAATAATACATAGAGAAATCCCTTACCTACAGAAACAAACACCTC 462  
 Y 344 SerGlyLeu 346  
 b 483 TCTGGTTTA 491  
 RESULT 10  
 ID ABA37272 standard; DNA; 494 BP.  
 IC ABA37272;  
 KC ABA37272;  
 KT 23-JAN-2002 (first entry)  
 CX Probe #15738 for gene expression analysis in human heart cell sample.  
 DE Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KM congenital heart disease; ss.  
 KV  
 KS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001MO-US000666.  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-0063236P.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 PT WPI; 2001-48899/53.  
 PT Single exon nucleic acid probes for analyzing gene expression in human hearts.  
 PS Claim 4; SEQ ID NO 15738; 530pp; English.  
 XX The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart. The  
 CC present sequence is one such probe. The probes may be used for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from the human heart via microarrays. By measuring gene expression, the  
 CC probes are useful for predicting, diagnosing, grading, staging,  
 CC monitoring and prognosing diseases of the human heart and vascular system  
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease. Note: The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
 Alignment Scores:  
 Pred. No.: 2,45e-84 Length: 494  
 Score: 886.00 Matches: 163  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 45.98% Indels: 0  
 DB: 4 Gaps: 0  
 US-10-010-050a-2 (1-346) x ABA37272 (1-494)  
 QY 184 AlAlvEtrPVallysgInAspAsnGluThrGlyIleTyrTyrGluThrTPAsnValLys 203  
 DB 3 GCAAGGGGTGAACAAGCAATGAAACAGAAATTATATATGACATGAAATGTAATA 62  
 QY 204 AlAserProGluLysGlyAlaGluThrTyrPheAspSerTyrAspCysSerLysPheVal 223  
 DB 63 GCCAGCCCAAGAAAAAGGGGCGAGAGACATGGTTGATCTTACGACCTTCCAAATTTGTC 122  
 QY 224 LeuArgThrPheAsnLysLeuAlaGluPheGlyAlaGluPheLysAsnIleGluThrAsn 243  
 DB 123 TTAAAGACCTTTAACAGAGTTGGCTGAATTTGAGCAGAGTTCAAGAACATGAAACCAAC 182  
 QY 244 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrLeuGlyAsnGluThrSerVal 263  
 DB 183 TATACAGAAATATTTCTTACAGTGAAGAACTTATCTGAGAAATGAAACATCTGTT 242  
 QY 264 PheGlyProThrGlyAsnLysThrLeuGlyLeuAlaIleLysAspPheTyrTyrProPhe 283  
 DB 243 TTTGGGCCAACAGAAACAAAGACTCTGTGTAGCCATTAAGATTTATTAACCCCTTC 302  
 QY 284 LysProHisLeuProThrLysGluPheLeuLeuSerLeuLeuGlnIlePheAspAlaVal 303  
 DB 303 AAACCAATTGGCCAACTAAAGAAATTTCTGTGAGTCTCTTGCAATTTTGATGCAGTG 362  
 QY 304 IleValHisLysGlnPheTyrLeuPheTyrAsnPheGlnTyrTyrPheLeuProMetLys 323  
 DB 363 ATTGTGCACAAACAGTCTCTATTGTTTATATATTTGAATTTGTTTAACTATGAAAA 422  
 QY 324 PheProPheIleLysIleThrTyrGluGlnIleProLeuProIleAspAsnLysThrLeu 343  
 DB 423 TTCCCTTTTATTAATAATACATAGAGAAATCCCTTACCTACAGAAACAAACACCTC 482  
 QY 344 SerGlyLeu 346  
 DB 483 TCTGGTTTA 491  
 RESULT 11  
 ID AAK44992 standard; DNA; 494 BP.  
 AC AAK44992;  
 AC AAK44992;  
 DT 06-NOV-2001 (first entry)  
 DE Human bone marrow expressed single exon probe SEQ ID NO: 19549.  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KM microarray; cancer; leukemia; lymphoma; myeloma; ss.  
 KV  
 KS Homo sapiens.  
 XX  
 PN WO200157276-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001MO-US000666.  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-0063236P.





XX WO200186003-A2.  
 XX 15-NOV-2001.  
 XX 30-JAN-2001; 2001WO-US000665.  
 XX 04-FEB-2000; 2000US-0180312P.  
 XX 26-MAY-2000; 2000US-0207456P.  
 XX 30-JUN-2000; 2000US-00608408.  
 XX 03-AUG-2000; 2000US-00632366.  
 XX 21-SEP-2000; 2000US-0234687P.  
 XX 27-SEP-2000; 2000US-0236359P.  
 XX 04-OCT-2000; 2000GB-00024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 XX measure gene expression in human lung samples.  
 XX Claim 4; SEQ ID NO 19228; 634bp; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 XX nucleic acid probes for measuring gene expression in a sample derived  
 XX from human lung comprising single exon nucleic acid probes having one of  
 XX 12614 nucleic acid sequences mentioned in the specification, or their  
 XX complements or the 12387 open reading frames derived from the 12614  
 XX probes. Also included are a microarray comprising the novel set of probes  
 XX; the novel set of probes which hybridize at high stringency to a nucleic  
 XX acid expressed in the human lung; measuring gene expression in a sample  
 XX derived from human lung, comprising (a) contacting the array with a  
 XX collection of detectably labeled nucleic acids derived from human lung  
 XX mRNA, and (b) measuring the label detectably bound to each probe of the  
 XX array; identifying exons in a eukaryotic genome, comprising (a)  
 XX algorithmically predicting at least one exon from genomic sequences of  
 XX the eukaryote; and (b) detecting specific hybridisation of detectably  
 XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe.  
 XX having a fragment identical to the predicted exon, the probe is included  
 XX in the above mentioned microarray; assigning exons to a single gene,  
 XX comprising (a) identifying exons from genomic sequence by the method  
 XX above and (b) measuring the expression of each of the exons in several  
 XX tissues and/or cell types using hybridisation to a single exon  
 XX microarrays having a probe with the exon, where a common pattern of  
 XX expression of the exons in the tissues and/or cell types indicates that  
 XX the exons should be assigned to a single gene; a peptide comprising one  
 XX of 12011 sequences, mentioned in the specification, or encoded by the  
 XX probes/open reading frames (ORF). The probes are used for gene expression  
 XX analysis; and for identifying exons in a gene, particularly using human  
 XX lung derived mRNA and for the study of lung diseases such as asthma, lung  
 XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 XX tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 XX Pudiak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 XX Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 XX present sequence is a single exon probe open reading frame of the  
 XX invention. Note: The sequence data for this patent did not form part of  
 XX the printed specification, but was obtained in electronic format directly  
 XX from WIPO at ftp.wipo.int/pub/published\_sequences  
 XX  
 XX SQ Sequence 494 BP; 170 A; 82 C; 90 G; 152 T; 0 U; 0 Other;  
 XX  
 XX Alignment Scores:  
 XX Pred. No.: 2.45e-84 Length: 494  
 XX Score: 886.00 Matches: 163  
 XX Percent Similarity: 100.00% Conservative: 0  
 XX Best Local Similarity: 100.00% Mismatches: 0  
 XX Query Match: 45.98% Indels: 0

DB: 6 Gaps: 0  
 US-10-010-050A-2 (1-346) x ABS19237 (1-494)  
 QY 184 AAlaYrTPVallysglnAspAsnGluThrGlyIleTyrTrpGluThrTrpAsnVallys 203  
 Db 3 GCNAAGTGGGTGAACAACGACCAATGAACAGAAATTTATATGTGACATGGAAATGTAATA 62  
 QY 204 AAlaSerProGluYulysGlyAlaGluThrTrpPheAspSerTyrAspCysSerTysPheVal 223  
 Db 63 GCCAGCCCAAGAAAGGGGGGACAGACATGTTGATTCCTACGACTGTTCCTCAAAATTTGTG 122  
 QY 224 LeuArgThrPheAsnYlsglAlaGluPheGlyAlaGluPheYlsglIleGluThrAsn 243  
 Db 123 TTAAGACCTTTACAAAGTTGGCTGAATTGGAGACAGATTCAAGACATGAAACCAAC 182  
 QY 244 TyrThrArgIlePheLeuTyrSerGlyGluProThrTyrIleGluYlsglIleThrSerVal 263  
 Db 183 TATACAAAGAAATATTTCTTATACATGAGAACCTTACTTCTGGGAAATGAACATCTGTT 242  
 QY 264 PheGlyProThrGlyAsnYlsglThrIleGluYlsglAlaIleYlsglAspPheTyrTrpPhe 283  
 Db 243 TTGGGGCAACGAAACGAAACAGACTTGTGTTAGCCATAAAGATTTATATACCCCTTC 302  
 QY 284 LysProHisLeuProThrIlysglIlePheLeuSerIleLeuGlnIlePheAspAlaVal 303  
 Db 303 AAACCACTTTGCCAATGAATTTCTGTTAGCTCTTGCACAAATTTTATGACAGTG 362  
 QY 304 IleValHisYlsglIlePheTyrIlePheTyrAsnPheGluTyrTrpPheLeuProMetIys 323  
 Db 363 ATGTGGCCAAACAGATTCATTTGTTTAAATTTGAATATGTTGTTTACTATGAAA 422  
 QY 324 PheProPheIleYlsglIleThrTyrGluGluIleProLeuProIleArgAsnYlsglIle 343  
 Db 423 TTCCTTTTATTAATTAATCAATGAGAAATCCCTTACCTATCGAAACAAACACATC 482  
 QY 344 SerGlyLeu 346  
 Db 483 TCTGGTTTA 491  
 Db  
 RESULT 15  
 AAV88347  
 ID AAV88347 standard; cDNA; 506 BP.  
 XX  
 XX AAV88347;  
 XX  
 XX 12-FEB-1999 (first entry)  
 XX  
 XX EST clone GB814.  
 XX  
 XX DE  
 XX  
 XX KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 XX KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 XX  
 XX OS Homo sapiens.  
 XX  
 XX XX  
 XX PN W0984537-A2.  
 XX  
 XX 15-OCT-1998.  
 XX  
 XX PD 10-APR-1998; 98WO-US006956.  
 XX  
 XX PF 10-APR-1997; 97US-00837312.  
 XX  
 XX PR 10-APR-1997;  
 XX  
 XX PA (GENY) GENETICS INST INC.  
 XX  
 XX PI Jacobs K, McCoy JM, Lavalie ER, Racie LA, Merberg D, Treacy M;  
 PI Spaulding V, Agostino MJ;  
 XX  
 XX WPI; 1999-070078/06.  
 XX  
 XX New polynucleotides encoding human secreted proteins - derived from e.g.

human blood, kidney, foetal lung, placenta, testes, brain, ovary,  
pituitary, retina and colon cDNA libraries.

Claim 1; Page 366; 6A1pp; English.

The present sequence represents an expressed sequence tag (EST), and is a polynucleotide of the invention. The polynucleotides of the invention are all secreted EST sequences isolated from a variety of human tissue sources. The EST sequences and proteins encoded by them are predicted to have useful biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals, although no supporting data is given. Suggested activities include nutritional activity, immune stimulating or suppressing activity, haematopoiesis regulating activity, tissue growth activity, haemostatic activity/inhibin regulating activity, chemotactic/chemokinetic activity, and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, tumour inhibition activity. The EST sequences are also stated to be useful for gene therapy

Sequence 506 BP; 171 A; 86 C; 119 G; 130 T; 0 U; 0 Other;

Alignment Scores:

Seq. No.:	1.85e-67	Length:	506
Score:	727.00	Matches:	133
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	37.73%	Indels:	0
DB:	2	Gaps:	0

US-10-010-050A-2 (1-346) x AAV88347 (1-506)

151	AlaAlaCysPhePheGluGlyIleAspAspValHisTrpLysGluAsnGlyThrLeuVal	170
99	GCTGCCCTCTTTTGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	158
171	GlnValAlaThrIleSerGlyAsnMetPheAsnGlnMetAlaLysTrpValLysGlnAsp	190
159	CAAGTAGCACTATATCATGAAACATGTTCAACCAAAATGGCAAGTGGTGAACAGGAC	208
191	AsnGluThrGlyIleTrpGluThrTrpAsnValLysAlaSerProGluLysGlyAla	220
219	AATGAAACAGAAATTATTATGACATGATGATGATGATGATGATGATGATGATGATGATGAT	278
211	GluThrTrpPheAspSerTrpAspCysSerLysPheValLeuArgThrPheAsnLysLeu	230
279	GAGCATGTGTTGATTCCTACGACTGTTCCAAATTGTGTTAAGGACCTTTAACAGTTG	338
231	AlaGluPheGlyAlaGluPheLysAsnIleGluThrAsnTrpThrArgIlePheLeuTrp	250
339	GCTGAATTTGGAGCAGAGTTCAAGACATGAAACCAACTATACAGAAATATTCTTTAC	398
251	SerGlyGluProThrTrpLeuGlyAsnGluThrSerValPheGlyProThrGlyAsnLys	270
399	AGTGAAGAACCTACTATCTGGGAAATGAAACATCTGTTTGGGCCAACGAAACAAG	458
271	ThrLeuGlyLeuAlaIleLysArgPheTrpProPhe	283
459	ACTCTTGTTTGGCATATAAAAGATTTTATACCCCTTC	497

Search completed: April 24, 2004, 23:26:22  
Job time : 646.658 secs





http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D1030CE010P1.

## FEATURES

source

Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1030Y01"  
/issue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-cligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 86.6%; Score 960.2; DB 9; Length 1201;  
Best Local Similarity 94.0%; Pred. No. 1.2e-199; Indels 8; Gaps 3;  
Matches 1012; Conservative 17; Mismatches 40;

```

QY      8 GCACGAGGCGAGGAGGTAGACACGCGACAGGGCGCGAGATGCGCGCGCGCGCGCG 67
DB      24 GCAGGCTGTGATCCGGTCCGGGATTCGCGGGATCGCGAGATGCGCGCGCGCGCGCG 63
QY      68 GCTGCGGAGCGCGCTTCTGCTGCTGCTGCGCGCTGCGCTGCTGCTGCTGCTGCTGCTG 127
DB      84 GCTGCGGAGCGCGCTTCTGCTGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143
QY      128 GGTGCTGCGCGCGGCTGCTGCGCGATCCCTCCGCGCGCACTGCGCGCTGCTGCTGCTGCTG 187
DB      144 GGTGCTGCGCGCGGCTGCTGCGCGATCCCTCCGCGCGCACTGCGCGCTGCTGCTGCTGCTG 203
QY      188 TTGACTTCGCTCCAAAACCTGATCTCTTATGTCAGCTAAGTACTTCTGCTCCAACT 247
DB      204 TTGACTTCGCTCCAAAACCTGATCTCTTATGTCAGCTAAGTACTTCTGCTCCAACT 263
QY      248 GGTGCTGCTCCAAAACCTGATCTCTTATGTCAGCTAAGTACTTCTGCTCCAACT 307
DB      264 GGTGCTGCTCCAAAACCTGATCTCTTATGTCAGCTAAGTACTTCTGCTCCAACT 323
QY      308 CCACTAATGGAATTAATATGAGACCTCTGCGGACACTGGAATATATGATGATGCC 367
DB      324 CCACTAATGGAATTAATATGAGACCTCTGCGGACACTGGAATATATGATGATGCC 383
QY      368 ATTGATTCAGAGATGATTAATGAGACCTCTGCGGACACTGGAATATATGATGATGCC 427
DB      384 ATTGATTCAGAGATGATTAATGAGACCTCTGCGGACACTGGAATATATGATGATGCC 443
QY      428 CAATGCGCACTGATCACTTCCCACTCCGACCTGGAATGAGATGCCCTTCTGCTGT 487
DB      444 CAATGCGCACTGATCACTTCCCACTCCGACCTGGAATGAGATGCCCTTCTGCTGT 503
QY      488 AATCAAGGCGCTGCTGCTTCTTGAAGGAAATGATGATGCTCACTGAGAGAGAAATGGG 547
DB      504 AATCAAGGCGCTGCTGCTTCTTGAAGGAAATGATGATGCTCACTGAGAGAGAAATGGG 563
QY      548 ACATTAAGTCAAGTGAACATATATGAGAAACATGTTCAACCAATATGCGAAAGTGGTG 607
DB      564 ACATTAAGTCAAGTGAACATATATGAGAAACATGTTCAACCAATATGCGAAAGTGGTG 623
QY      608 AAACAGGAGCAATGAACAGGAATTTATATGAGACATGGAATGTAAGAACAGCCAGCCAGAA 667
DB      624 AAACAGGAGCAATGAACAGGAATTTATATGAGACATGGAATGTAAGAACAGCCAGAA 683
QY      668 AAGGGGCGAGAGCATGTTGATTCCTACGACTGTTCCAAATTTGTGTTAAGACCTTT 727
DB      684 AAGGGGCGAGAGCATGTTGATTCCTACGACTGTTCCAAATTTGTGTTAAGACCTTT 743
QY      728 AACAGTTGCTGAATTTGAGAGAGAGTTCAAGAAATATGAACCAATATCAAGAAATA 787
DB      744 AACAGTTGCTGAATTTGAGAGAGAGTTCAAGAAATATGAACCAATATCAAGAAATA 803
QY      788 TTCTTTAAGTGAAGAACTACTATCTGGAAGAAATGAACATCTGTTTGTGGGCCAACA 847

```

```

DB      804 TTCTTTAAGTGAAGAACTACTATCTGGAAGAAATGAACATCTGTTTGTGGGCCAACA 863
QY      848 GGAACAGAGCTTGTGTTAGCCATAAAGATTTATATACCTTCAACCAATTTG 907
DB      864 GGAACAGAGCTTGTGTTAGCCATAAAGATTTATATACCTTCAACCAATTTG 923
QY      908 CCACTAATGGAATTAATATGAGACCTCTGCGGACACTGGAATATATGATGATGCC 967
DB      924 CCACTAATGGAATTAATATGAGACCTCTGCGGACACTGGAATATATGATGATGCC 983
QY      968 CAGTTCTATTTGTTTATATATGAGATGATGATGATGATGATGATGATGATGATGATG 1027
DB      984 CAGTTCTATTTGTTTATATATGAGATGATGATGATGATGATGATGATGATGATGATG 1038
QY      1028 AAATATACATATGAAGAAATCCCTTACTATACAGAAACAAACACTCTGCTTTA 1084
DB      1039 TAAATACATATGAAGAAATCCCTTACTATACAGAAACAAACACTCTGCTTTA 1092

```

## RESULT 2

BX331615

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 923)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DB009AB010P1&cluster=7238.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DB009AB010P1.

## FEATURES

source

Location/Qualifiers  
1. 923  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DB009Y01"  
/issue\_type="NEUROBLASTOMA COT 10-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-cligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 74.9%; Score 811.6; DB 13; Length 923;  
Best Local Similarity 96.8%; Pred. No. 4.2e-167; Indels 0; Gaps 0;  
Matches 811; Conservative 9; Mismatches 18;

```

QY      14 GGCAGGAGGTGAGACAGGCGCCGAGATGCGCGCGCGCGCGCGCGCGCTGG 73
DB      85 GGCAGGAGGTGAGACAGGCGCCGAGATGCGCGCGCGCGCGCGCGCGCTGG 144
QY      74 GGAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 133
DB      145 GGAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 204

```

	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

```

FEATURES
  source
    1. 1077
    2. 1077
    3. 1077
    4. 1077
    5. 1077
    6. 1077
    7. 1077
    8. 1077
    9. 1077
    10. 1077
    11. 1077
    12. 1077
    13. 1077
    14. 1077
    15. 1077
    16. 1077
    17. 1077
    18. 1077
    19. 1077
    20. 1077
    21. 1077
    22. 1077
    23. 1077
    24. 1077
    25. 1077
    26. 1077
    27. 1077
    28. 1077
    29. 1077
    30. 1077
    31. 1077
    32. 1077
    33. 1077
    34. 1077
    35. 1077
    36. 1077
    37. 1077
    38. 1077
    39. 1077
    40. 1077
    41. 1077
    42. 1077
    43. 1077
    44. 1077
    45. 1077
    46. 1077
    47. 1077
    48. 1077
    49. 1077
    50. 1077
    51. 1077
    52. 1077
    53. 1077
    54. 1077
    55. 1077
    56. 1077
    57. 1077
    58. 1077
    59. 1077
    60. 1077
    61. 1077
    62. 1077
    63. 1077
    64. 1077
    65. 1077
    66. 1077
    67. 1077
    68. 1077
    69. 1077
    70. 1077
    71. 1077
    72. 1077
    73. 1077
    74. 1077
    75. 1077
    76. 1077
    77. 1077
    78. 1077
    79. 1077
    80. 1077
    81. 1077
    82. 1077
    83. 1077
    84. 1077
    85. 1077
    86. 1077
    87. 1077
    88. 1077
    89. 1077
    90. 1077
    91. 1077
    92. 1077
    93. 1077
    94. 1077
    95. 1077
    96. 1077
    97. 1077
    98. 1077
    99. 1077
    100. 1077
    101. 1077
    102. 1077
    103. 1077
    104. 1077
    105. 1077
    106. 1077
    107. 1077
    108. 1077
    109. 1077
    110. 1077
    111. 1077
    112. 1077
    113. 1077
    114. 1077
    115. 1077
    116. 1077
    117. 1077
    118. 1077
    119. 1077
    120. 1077
    121. 1077
    122. 1077
    123. 1077
    124. 1077
    125. 1077
    126. 1077
    127. 1077
    128. 1077
    129. 1077
    130. 1077
    131. 1077
    132. 1077
    133. 1077
    134. 1077
    135. 1077
    136. 1077
    137. 1077
    138. 1077
    139. 1077
    140. 1077
    141. 1077
    142. 1077
    143. 1077
    144. 1077
    145. 1077
    146. 1077
    147. 1077
    148. 1077
    149. 1077
    150. 1077
    151. 1077
    152. 1077
    153. 1077
    154. 1077
    155. 1077
    156. 1077
    157. 1077
    158. 1077
    159. 1077
    160. 1077
    161. 1077
    162. 1077
    163. 1077
    164. 1077
    165. 1077
    166. 1077
    167. 1077
    168. 1077
    169. 1077
    170. 1077
    171. 1077
    172. 1077
    173. 1077
    174. 1077
    175. 1077
    176. 1077
    177. 1077
    178. 1077
    179. 1077
    180. 1077
    181. 1077
    182. 1077
    183. 1077
    184. 1077
    185. 1077
    186. 1077
    187. 1077
    188. 1077
    189. 1077
    190. 1077
    191. 1077
    192. 1077
    193. 1077
    194. 1077
    195. 1077
    196. 1077
    197. 1077
    198. 1077
    199. 1077
    200. 1077
    201. 1077
    202. 1077
    203. 1077
    204. 1077
    205. 1077
    206. 1077
    207. 1077
    208. 1077
    209. 1077
    210. 1077
    211. 1077
    212. 1077
    213. 1077
    214. 1077
    215. 1077
    216. 1077
    217. 1077
    218. 1077
    219. 1077
    220. 1077
    221. 1077
    222. 1077
    223. 1077
    224. 1077
    225. 1077
    226. 1077
    227. 1077
    228. 1077
    229. 1077
    230. 1077
    231. 1077
    232. 1077
    233. 1077
    234. 1077
    235. 1077
    236. 1077
    237. 1077
    238. 1077
    239. 1077
    240. 1077
    241. 1077
    242. 1077
    243. 1077
    244. 1077
    245. 1077
    246. 1077
    247. 1077
    248. 1077
    249. 1077
    250. 1077
    251. 1077
    252. 1077
    253. 1077
    254. 1077
    255. 1077
    256. 1077
    257. 1077
    258. 1077
    259. 1077
    260. 1077
    261. 1077
    262. 1077
    263. 1077
    264. 1077
    265. 1077
    266. 1077
    267. 1077
    268. 1077
    269. 1077
    270. 1077
    271. 1077
    272. 1077
    273. 1077
    274. 1077
    275. 1077
    276. 1077
    277. 1077
    278. 1077
    279. 1077
    280. 1077
    281. 1077
    282. 1077
    283. 1077
    284. 1077
    285. 1077
    286. 1077
    287. 1077
    288. 1077
    289. 1077
    290. 1077
    291. 1077
    292. 1077
    293. 1077
    294. 1077
    295. 1077
    296. 1077
    297. 1077
    298. 1077
    299. 1077
    300. 1077
    301. 1077
    302. 1077
    303. 1077
    304. 1077
    305. 1077
    306. 1077
    307. 1077
    308. 1077
    309. 1077
    310. 1077
    311. 1077
    312. 1077
    313. 1077
    314. 1077
    315. 1077
    316. 1077
    317. 1077
    318. 1077
    319. 1077
    320. 1077
    321. 1077
    322. 1077
    323. 1077
    324. 1077
    325. 1077
    326. 1077
    327. 1077
    328. 1077
    329. 1077
    330. 1077
    331. 1077
    332. 1077
    333. 1077
    334. 1077
    335. 1077
    336. 1077
    337. 1077
    338. 1077
    339. 1077
    340. 1077
    341. 1077
    342. 1077
    343. 1077
    344. 1077
    345. 1077
    346. 1077
    347. 1077
    348. 1077
    349. 1077
    350. 
```

ORIGIN

Query Match	73.1%;	Score 792.4;	DB 13;	Length 1011;
Best Local Similarity	90.0%;	Pred. No. 6.7e-163;		
Matches 821;	Conservative 49;	Mismatches 36;	Indels 6;	Gaps 3;

[illegible]

733 GTTGCTGAATTGGAGCAGGTTCAAGAACATAGAAACCACTATACAGAAATATTTCT 752  
 802 ATTGCTGAATTTTGGAGAGGTTCAAAACAAAACCAATATACAAAATATTTCT 861  
 793 TTACAGTGAAGACCTACTATCTGGAAATGATAACATCTGTTTGGGCCCAAGAAA 852  
 862 TTWAGTGGKAR-CTACTTATCTKGGGAATGAAACATCTTTTGGGACARGGAC 920  
 853 CAAGACTCTGTTAGCCATPAAAAAGATTATATCCCTTCAACCACTTTGCCAAC 922  
 921 MAACCTCTTGCTTGGCMTAAAGATTWTTA-----SCYTCMARCCMTTSCMHA 976  
 913 TAAAGATTTCT 924  
 977 AAAATKYTTTT 988

RESULT 4  
 LOCUS B1090566  
 DEFINITION 602855673P1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:4996891 5',  
 mRNA sequence.

ACCESSION B1090566 869 bp mRNA linear EST 20-JUN-2001  
 VERSION B1090566  
 KEYWORDS B1090566.1 GI:14508896  
 EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 869)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLM at:  
 http://image.llnl.gov  
 Plate: LHM1023 row: a column: 20  
 High quality sequence stop: 843.

FEATURES  
 source  
 1..869  
 location/Qualifiers

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4996891"  
 /cell\_line="MGC36"  
 /lab\_host="DH10B"  
 /clone\_id="NIH\_MGC\_10"  
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: Not; Site 2: SalI; Cloned unidirectional. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."

## ORIGIN

Query Match 69.1%; Score 749.4; DB 12; Length 869;  
 Best Local Similarity 97.1%; Pred. No. 1.7e-153;  
 Matches 827; Conservative 0; Mismatches 17; Indels 8; Gaps 6;  
 14 GGGCAGAGTAGACACGCGACAGGCGCGGAGATGGCGCGGGCGCGGGCGCGGCTCGG 73  
 19 GCGCAGGAGTAGACACGCGACAGGCGCGCGAGATGGCGCGGGCGCGGGCGCGGCTCGG 78  
 74 GAGCGCGCTTCTGCTGCTGGCGCGCTGGCGCTGCTTGGCTCGCGGTGTTCCGGGCTGG 133  
 79 GAGCGCGCTTCTGCTGCTGGCGCGCTGGCGCTGCTTGGCTCGCGGTGTTCCGGGCTGG 138  
 134 TCCCGGCTCTCGGGCATCCCTCCCGGGCGCACTGGCGGTGCCCTACAGCGCTTTGAC 193

Db 139 TCCCGGCTCTCGGGCATCCCTCCCGGGCGCACTGGCGGTGCCCTACAGCGCTTTGAC 198  
 194 TTCCGTCAAAACCTGATCCTTATTTGTCAAGCTAAGATATCTTCTGTCACTGCTCA 253  
 199 TTCCGTCAAAACCTGATCCTTATTTGTCAAGCTAAGATATCTTCTGTCACTGCTCA 258  
 254 CCTATCCAGTTATGAGGGGTGATGATGACATTGAACTTTTGGATTACAAGCCCCAGTA 313  
 259 CCTATCCAGTTATGAGGGGTGATGATGACATTGAACTTTTGGATTACAAGCCCCAGTA 318  
 314 TGGGAATTTAAATATGAGACCTTCCTGGGACCTTGGAAATTAATGATGATGATGGA 373  
 319 TGGGAATTTAAATATGAGACCTTCCTGGGACCTTGGAAATTAATGATGATGATGGA 378  
 374 TTGGAAGTACATTAATCTGCAAGAACTACACAAATGAAATGAAATGAAATGAAATGAA 433  
 379 TTGGAAGTACATTAATCTGCAAGAACTACACAAATGAAATGAAATGAAATGAAATGAA 438  
 434 GGCACATGTAATTTCCCATCTTCGACCTGGAATGAAATGAAATGAAATGAAATGAAATGAA 493  
 439 GGCACATGTAATTTCCCATCTTCGACCTGGAATGAAATGAAATGAAATGAAATGAAATGAA 498  
 494 GGCCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 553  
 499 GGCCTGCTGCTTTTGGAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558  
 554 GTTCAGTAGCACTATATACGAAACATGTTCAACCAATGCGCAATGGTGGTGAACAG 613  
 559 GTTCAGTAGCACTATATACGAAACATGTTCAACCAATGCGCAATGGTGGTGAACAG 617  
 614 GACATGTAACAGAAATTTATATGAGCATGGAATGTAAGCAAGCCAGCCAGAA--AAGG 671  
 618 GACATGTAACAGAAATTTATATGAGCATGGAATGTAAGCAAGCCAGCCAGAAAGCC 677  
 672 GGGCAGAGACATGTTTATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 731  
 678 GGCACGAGCATGTTTATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 737  
 732 AGTTGGCTGAA-TTGGAGCAGAGTTCAAGAA-TAGAAACCACTATACGAATAT 789  
 738 AGTTGGCTGAACTTGGAGCAGAGTTCAAGAACTAGAAACCACTATACGAATAT 797  
 790 TCTTTACAGTGG-AACAACCTATCTTGGGAAA--TGAAACATCTGTTTGGGCGAAC 846  
 798 TCTTTACAGTGGAAACCTATCTTGGGAAAATGAAACATCTGTTTGGGCGAAC 857  
 847 AGGAAACAGAC 858  
 858 AGGAAACAGAC 869

RESULT 5  
 CA488543 924 bp mRNA linear EST 14-NOV-2002  
 LOCUS AGENCOURT\_10808864 MAPCL Homo sapiens cDNA clone IMAGE:6720309 5',  
 DEFINITION mRNA sequence.

ACCESSION CA488543  
 VERSION CA488543  
 KEYWORDS CA488543.1 GI:24950702  
 EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 924)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cgabs-remail.nih.gov  
 Tissue Procurement: Kristi A. Eglund, Ira Pastan  
 cDNA Library Preparation: Invitrogen Corp.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation





y		954	AAGCTCTGGTTTACCATTAAGAATTT	884
b		830	CCGCCCTCGGTCCTTCTCCTCAAGGAAATT	860
RESULT 8 K043247				
OCUS EFINITION				
AK043247		2356 bp	mRNA	linear HTC 19-SEP-2003
Mus musculus 7 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:A730075N08 product:similar to CERDID-LIPPEUSCUSINIS NEURONAL PROTEIN 5 (CLNS PROTEIN) [Homo sapiens], full insert sequence.				
ACCESSION ERSION KEYWORDS SOURCE ORGANISM				
AK043247.1		GI:26089577		
HTC; CAP trapper. Mus musculus (house mouse) Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PubMed				
Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)				
99279253 10349636				
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PubMed				
Carninci, P., Shibata, Y., Hayatsu, N., Sugihara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)				
20499374 11042159				
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PubMed				
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Niemi, K., Kitamura, T., Tashtiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasaiwaagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, I., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)				
11076861				
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PubMed				
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)				
5				
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PubMed				
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)				
6 (bases 1 to 2356)				
ADACHI, J., AIZAWA, K., AKIMURA, T., ARAKAWA, T., BONO, H., CARNINCI, P., FUJUKAWA, S., FURUNO, M., HANAGAKI, T., HARA, A., HASHIZUME, M., HAYASHIDA, K., HAYATSU, N., HIMOTO, K., HIROCKA, T., HITOZANE, T., HOEI, F., IMOTANI, K., ISHII, Y., ITOH, M., KAGAWA, I., KASUKAWA, T., KATO, H., KAWAI, J., KOJIMA, Y., KONDO, S., KOMNO, H., KONDA, M., KOYO, S., KURIHARA, C., MATSUMURA, T., MIYAZAKI, A., MURATA, M., NAKAMURA, M., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OHSATO, N., OKAZAKI, Y., SAITO, R., SATOH, H., SAKAI, C., SAKAI, C., SAKAZUME, N., SANO, H., SASAKI, D., SHIBATA, K., SHINGAWA, A., SHIRAKI, T., SOGABE, Y., TAGAMI, M., TAGAWA, A., TAKAHASHI, F., TAKAKU-AKAHIRA, S., TAKEDA, Y., TANAKA, T., TOMARI, A., TOYA, T., TSUMISHI, A., MURAMATSU, M. and HAYASHIZAKI, Y.				
Direct Submission Submitted (16-JUN-2001) Yoshitake Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome				

COMMENT

Exploitation Research Group RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

FEATURES

source

1. .2356  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:A730075N08"  
/db\_xref="MG1:2408405"  
/db\_xref="taxon:10090"  
/clone="A730075N08"  
/tissue\_type="cerebellum"  
/clone\_id="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="7 days neonate"

misc\_feature

1. .2356  
/note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5  
(CNS5 PROTEIN) [Homo sapiens] (SWISSPROT O75503, evidence:  
FASTA, 73.9%ID, 86.7%length, match=1041)"

ORIGIN

Query Match 62.7%; Score 680.2; DB 11; Length 2356;  
Best Local Similarity 79.3%; Pred. No. 2.5e-138;  
Matches 821; Conservative 0; Mismatches 208; Indels 6; Gaps 1;

Db 49 GCGGCGGGGCGCGGCGCGCTCGGGGACGCGCTTCCTGCGTGGCGCTGGCGCTGCT 108  
48 GCGCGGAGATGCTGCGCGGCGGGGCATGTGAGCGCAGCTGGCGCCCGCGGCTCGCGTGGC 107  
109 TTGGCTCGCGGCGTTCGCGGCGCGCTCCGGGCTCGGGGCATCCCGCTCGGCGCATCG 168  
108 GCTGCTCGGCGTGGGAC-----GATTCTGGCGCGCTCCCGAGCTTGGCGCAAGCGTG 161  
169 GCGGCGGCGCTACAGCGCTTGACTTCGCTCCGTCGCAAACTGATCTTATTTGCAAGCTAA 228  
162 GCGGCGGCGCTACAGCGCTTCTCTTCCGTCGGAAGACAGATCCCTACTGTCAAGCTAA 221  
229 GATACTTCTGTCCAATGCTGCTCACTATCCGATGATGAGGGTATGATGACATTTGA 288  
222 GTTACTTCTGTCTCTCAACGGCTCGCGCCATCCAGTTATGAAAGCAAGAGCTCATGCA 281  
289 AGTTTTCATTCAGAGCCCGCAATGGAATTTAAATTTGAGACCTCCTGGGACACTT 348  
282 GGTCTTACAGACTACAGCCCGCATTTGGGAATTTAAATTTGAGACTCTCTGGACACTT 341  
349 GAAATTTAGCATGATGCGCATTTGATTCAGAGTACATTAATCTGGCAAGACTTACAT 408  
342 TAAACTTATGCAAGACGCGTGGGATTCAGAGGACATCTGACAGCAAGCAATACAT 401  
409 GGAATGATGACTTTTCCAACTTGGCACTGTACATTTCCCATCTCCGACTTGAAT 468  
402 TGAATGTATGAATCTTTCCAGCTGGGCACTGTACATTTCCCACTCGGCTTGACAA 461  
469 GGAATGCCCTTTGTGTATATCAAGCGCTGCTCTTTTGAAGGAATGATGATG 528  
462 GAGGCGCTCCTTCTGGGTATCAAGGAGGCGAGCTGTATTTTGAAGGAATGATGAT 521  
529 TCATGGAAGAAATATGAGCATTTAGTTCAGTGAAGCACTATTCAGGAAACATGTTCA 588  
522 ACACTGGAAGAAACCGGACACTGTCACTGTTCACATATTCGGGAAGACATTTAA 581  
589 CCAATGGCAAGGCGGTGAAGCAAGCATATGAAACGAATTTATGAGCATGGA 648



JOURNAL  
Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE  
20530913

PubMed  
11076861

REFERENCE  
4

TITLE  
The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.

JOURNAL  
Functional annotation of a full-length mouse cDNA collection

REFERENCE  
5

TITLE  
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

JOURNAL  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

REFERENCE  
6

TITLE  
Nature 409, 685-690 (2001)

JOURNAL  
Nature 420, 563-573 (2002)

REFERENCE  
6 (bases 1 to 2433)

TITLE  
Aachi,U., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuo,M., Hamagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoch,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,U., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Ozaki,Y., Saito,R., Shibata,K., Shinagawa,A., Shiraki,T., Sano,H., Sasaki,D., Shibata,K., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,M., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

JOURNAL  
Submitted (16-APR-2002)

REFERENCE  
Yoshihide Hayashizaki, The Institute of Direct Submission

TITLE  
Exploitation and Chemical Research (RIKEN), Laboratory for Genome Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Seihori-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscl.riken.go.jp), Fax:81-45-503-9216

COMMENT  
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>  
Location/Qualifiers

source  
1..2433  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM\_DB:D730033P03"  
/db\_xref="MG1:2422937"  
/db\_xref="taxon:10090"  
/clone="D730033P03"  
/sex="female"  
/tissue\_type="mammary gland"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days lactation, adult"  
1..2433  
/note="similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CIN5 PROTEIN) [Homo sapiens] (SWISSPROT|O75503, evidence FASTA, 73.9%ID, 86.7%length, match=1041)"

ORIGIN  
Query Match 62.6%; Score 678.6; DB 11; length 2433;  
Best Local Similarity 79.2%; Pred. No. 5,7e-138;  
Matches 820; Conservative 0; Mismatches 209; Indels 6; Gaps 1

DB  
49 GCGCGGGGCGCGGGCGCGGCTTGGGAGGCGCTTCGTGCTGGGCGCTGGCGCTGCT 108  
125 GCGCGGATCTGCTGCGGGGCGCATGTGAGGCACTGGCCCGCGCTGCGCTGGCG 184  
109 TTGGCTGCGGATGCTTCCGGGCTGTCCGGGCTTCTCGGCGATCCCTCCGCGGCACTG 168  
185 GCTGCTCGGCTGCGCAC-----GATTCTGGGCGGCTCCGAGACGTCTGGGCAACGCTG 238



```

169 GCCGCTGCCCTACAGCGCTTGAATTCGTCACAAACCTGATTCCTATTGTCAAGCTAA 223
b 239 GCCGGTGGCCCTACAGCGCTTCTCTTCCGTCCGAAAGACAGATCCCTACGTCAACCTAA 293
Y 229 GTATACCTTCTGTCCAACTGCTCACCCTATCCAGTTATGAGAGGTGATGATCACTTGA 288
b 299 GTATACCTTCTGTCCAACTGCTCACCCTATCCAGTTATGAGAGGTGATGATCACTTGA 358
Y 289 AGTTTTCATTCACAGCGCTTGAATTCGTCACAAACCTGATTCCTATTGTCAAGCTAA 348
b 359 GGCTTTCACATCAAGCGCTTGAATTCGTCACAAACCTGATTCCTATTGTCAAGCTAA 418
Y 349 GAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
b 419 TAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 478
Y 409 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 468
b 479 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 538
Y 469 GGATGCGCCCTTCTGTGATTAATCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
b 539 GAGCGCTCCCTTCTGTGATTAATCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598
Y 529 TCATCTGGAAGAAATGGGACATTAATCAAGTACCACTATATCAGGAAACATGTTCAA 538
b 599 ACATCTGGAAGAAATGGGACATTAATCAAGTACCACTATATCAGGAAACATGTTCAA 658
Y 589 CCAATGGAAGAAATGGGACATTAATCAAGTACCACTATATCAGGAAACATGTTCAA 648
b 659 CAAATGGAAGAAATGGGACATTAATCAAGTACCACTATATCAGGAAACATGTTCAA 718
Y 649 TGTAAAAGCCAGCCCAAGAAAGGGGGCAGAGCATGTTGATTCCTACAGCTGTCCA 708
b 719 GGTCCAGCGCGCTCAGACCAAGGGGCCAGACGTGTTGAGTCTACAGCTGTGTGAA 778
Y 709 ATTTGTTTGAAGACCTTTTAAACAAGTGTGATTTGAGAGAGAGTTGAGAGAGAGTGA 768
b 779 TTTTGTCTTAAAGACATTAATGAAATTTGCTGATTTGGAACAGATTCAGAGAGATGA 838
Y 769 AACCACTATACCAAAATATTTCTTTACGTGAGAGCTTATTAATCTGGGAAATGAAAC 828
b 839 AACCACTATACCAAAATATTTCTTTACGTGAGAGCTTATTAATCTGGGAAATGAAAC 898
Y 829 ATCTGTTTGGGCAACAGAAACAGACTCTTGGTTTACGCTATTAATTAATTAATTA 888
b 899 ATCTGTTTGGGCAACAGAAACAGACTCTTGGTTTACGCTATTAATTAATTAATTA 958
Y 889 CCGCTTCAACACATTTGCACTAAGAAATTTGTTGAGTCTCTGCAAAATTTTGA 948
b 959 CCGCTTCAACACATTTGCACTAAGAAATTTGTTGAGTCTCTGCAAAATTTTGA 1018
Y 949 TGCAGTATGTCACAAACAGTCTATTTGTTTATTAATTTGAATTTGTTTACC 1008
b 999 TGCAGTATGTCACAAACAGTCTATTTGTTTATTAATTTGAATTTGTTTACC 1078
Y 1009 TATGAATTTCCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1068
b 1079 TATGAATTTCCCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1138
Y 1069 AACACTCTCTGTTT 1083
b 1139 AACATTTACCGACTT 1153

```

```

RESULT 10
AL571805/c 1201 bp mRNA linear EST 31-MAY-2003
LOCUS AL571805 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CS0D1030YJ01 3-PRIME, mRNA sequence.
ACCESSION AL571805
VERSION AL571805.2 GI:31293196

```

```

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 1201)
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
COMMENT On Feb 16, 2001 this sequence version replaced gi:12929467.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 7238.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1030CE01NP1&cluster=7238.f. Contact :
Peng Liang Email : filiang@life.techn.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1030CE01NP1.
Location/Qualifiers
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1030YJ01"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dt)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

```

## ORIGIN

```

Query Match 62.2%; Score 674.2; DB 9; Length 1201;
Best Local Similarity 97.8%; Pred. No. 5.1e-137;
Matches 704; Conservative 9; Mismatches 4; Indels 3; Gaps 3;
QY 366 CCATTGATTCAGAAATGATTAATCTGCAAGCAAGTCAATGATGATGATGATGATGATGAT 425
DB 1075 SCATGATTCAGAAATGATTAATCTGCAAGCAAGTCAATGATGATGATGATGATGATGAT 1018
QY 426 TCCAACTTGGCAACGTACATTTCCCATCTCCGACTCGAATGATGATGATGATGATGATGAT 485
DB 1017 TCCAACTTGGCAACGTACATTTCCCATCTCCGACTCGAATGATGATGATGATGATGATGAT 958
QY 486 GTATCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544
DB 957 GTATCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 898
QY 545 GGGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 604
DB 997 GGGACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 838
QY 897 GTGAAAGAGCAATGAAACAGAAATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 778
DB 837 GTGAAAGAGCAATGAAACAGAAATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 724
QY 605 GTGAAAGAGCAATGAAACAGAAATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 664
DB 665 GAAAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 724
QY 777 GAAAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 718
DB 725 TTTAACAAGTGGCTGATTTGGAGCAGAGTTCAAGAACATAGAAACCACTATACAGA 784
QY 717 TTTAACAAGTGGCTGATTTGGAGCAGAGTTCAAGAACATAGAAACCACTATACAGA 658
DB 785 ATATTTCTTTACAGTGAAGAACTACTATCTGGAAGAAATGAAACATCTGTTTGGGCA 844
QY 657 ATATTTCTTTACAGTGAAGAACTACTATCTGGAAGAAATGAAACATCTGTTTGGGCA 598
DB 845 ACAGAAACAGAACTCTGTTTACCCATTAATTAATTAATTAATTAATTAATTAATTAATTA 904

```

DB 597 ACAGAAACAGACCTGGTTAGCCATAAAATTTATTTACCCCTTAACACCAT 538  
 QY 905 TTGGCACTGAAGAAATTTCTGTAGTCTCTGCAATTTTGTATGACAGTGTGAC 934  
 DB 537 TTGGCACTGAAGAAATTTCTGTAGTCTCTGCAATTTTGTATGACAGTGTGAC 478  
 QY 965 AACAGTTCTATTTGTTTATTAATTTTGAATATGTTTATGCTATGAAATTCCTTT 1324  
 DB 477 AACAGTTCTATTTGTTTATTAATTTTGAATATGTTTATGCTATGAAATTCCTTT 418  
 QY 1025 ATTTAAATATACATATGAAAGAAATCCCTTACCTATCAGAAACAAACCTCTGCTTTA 1084  
 DB 417 ATTTAAATATACATATGAAAGAAATCCCTTACCTATCAGAAACAAACCTCTCTGCTTTA 358  
 RESULT 11  
 BE873363 931 bp mRNA linear EST 20-OCT-2000  
 LOCUS 601450425F1 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3854119 5',  
 DEFINITION mRNA sequence.  
 BE873363  
 ACCESSION BE873363.1 GI:10322139  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 931)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LHM9579 row: b column: 08  
 High quality sequence, stop: 662.  
 Location/Qualifiers  
 1..931  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3854119"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_id="NIH\_MGC\_65"  
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1; NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.8 kb. Library constructed by Life  
 Technologies."  
 ORIGIN  
 Query Match 62.1%; Score 673; DB 10; Length 931;  
 Best Local Similarity 93.0%; Pred. No. 9,3e-137;  
 Matches 727; Conservative 0; Mismatches 51; Indels 4; Gaps 2;  
 QY 274 TGATGATGACATGTAAGTTTTCGATTAACAAGCCCAAGTATGGAAATTAATATGAGA 333  
 DB 1 TGATGATGACATGTAAGTTTTCGATTAACAAGCCCAAGTATGGAAATTAATATGAGA 60  
 QY 334 CCTCTGGGACCTTGAATAATATGATGATGATGATGATGATGATGATGATGATGATG 393  
 DB 61 CCTCTGGGACCTTGAATAATATGATGATGATGATGATGATGATGATGATGATGATG 120  
 QY 394 CAAGAATCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 453  
 DB 121 CAAGAATCAACAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180

QY 454 TCTCCGACCTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 513  
 DB 181 TCTCCGACCTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
 QY 514 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 573  
 DB 241 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
 QY 574 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 633  
 DB 301 AGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360  
 QY 634 TTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 693  
 DB 361 TTATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
 QY 694 CTACGACTGTTCCAAATTTGTTTATGATGATGATGATGATGATGATGATGATGATGATG 753  
 DB 421 CTACGACTGTTCCAAATTTGTTTATGATGATGATGATGATGATGATGATGATGATGATG 480  
 QY 754 GTTCAAGAACATGAAACCACTATACAGAAATTTCTTTACAGTGGAGAACTACTTA 813  
 DB 481 GTTCAAGAACATGAAACCACTATACAGAAATTTCTTTACAGTGGAGAACTACTTA 540  
 QY 814 TCTGGGAATGAAACATCTGTTTGGGCAACAGGAAACAGACTCTTGTGTAGCCAT 873  
 DB 541 TCTGGGAATGAAACATCTGTTTGGGCAACAGGAAACAGACTCTTGTGTAGCCAT 600  
 QY 874 AAAAGATTTTATTTACCTCTTCAACCACTTTGCACTTATGCAATTTCTGTAGTCT 933  
 DB 601 AAAAGATTTTATTTACCTCTTCAACCACTTTGCACTTATGCAATTTCTGTAGTCT 660  
 QY 934 CTGGAATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 993  
 DB 661 CTGGAATTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716  
 QY 994 ATATTGTTTATTTACCTATGAAATTCCTTTTATTAATATGATGATGATGATGATGATG 1053  
 DB 717 ATATTGTTTATTTACCTATGAAATTCCTTTTATTAATATGATGATGATGATGATGATG 776  
 QY 1054 AC 1055  
 DB 777 TC 778  
 RESULT 12  
 BU621797/c 689 bp mRNA linear EST 23-SEP-2002  
 LOCUS BU621797  
 DEFINITION UI-H-FIL-bdb-1-07-0-UI.s1 NCI CGAP FIL Homo sapiens cDNA clone  
 BU621797  
 ACCESSION BU621797.1 GI:23288012  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 689)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: James Martin  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 Seq primer: M13 FORWARD  
 POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
1. 689  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-Flu-Bdp-1-07-0-UI"  
/tissue\_type="Cell lines"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_11b="NCI CGAP FL1"  
/note="Organ: Chondrosarcoma; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: Ecor I; Site 2: Not I; NCI CGAP FL1 is a normalized cDNA library derived from a pool of mRNA obtained from 4 cell lines from grade III chondrosarcoma tissues. The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GAGGTCCGTC. The cell lines were provided by Dr. James Martin from the University of Iowa.  
TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix  
TAG\_Lib=UI-H-Flu  
TAG\_SEQ=GAGGTCCGTC"

## JRGIN

Query Match 61.9%; Score 670.6; DB 13; Length 689;  
Best Local Similarity 99.4%; Pred. No. 3,1e-136;  
Matches 673; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

370 TGGATTGAGAGTACATTACTGCGAAGACTACAACTGAAATGGTATGACTTTTCCA 429  
369 TGGATTGAGAGTACATTACTGCGAAGACTACAACTGAAATGGTATGACTTTTCCA 630  
430 ACTTGGCACTGTTCATTTCCCACTCCGACCTGGAATGGATGCCCCCTTCTGGTGA 489  
629 ACTTGGCACTGTTCATTTCCCACTCCGACCTGGAATGGATGCCCCCTTCTGGTGA 570  
490 TCAAGGCGCTGCTCTTTTGAAGGAATGATGATGTTCACTGGAAGAAATGGGAC 549  
569 TCAAGGCGCTGCTCTTTTGAAGGAATGATGATGTTCACTGGAAGAAATGGGAC 510  
550 ATTAGTTCAAGTAGCACTATATTCAGGAACATGTTCAACCAATGGCAAAAGTGGTAA 639  
509 ATTAGTTCAAGTAGCACTATATTCAGGAACATGTTCAACCAATGGCAAAAGTGGTAA 430  
610 ACAGGACATGAAAGAGAAATTTATGAGACATGATGATGAAAGCCAGCCAGAAA 659  
449 ACAGGACATGAAAGAGAAATTTTGAAGCATGATGATGAAAGCCAGCCAGAAA 390  
670 GGGGGCAGAGCATGTTGATCTCTACGACTGTTCAAAATTTGTTGAAGACCTTTAA 729  
389 GGGGGCAGAGCATGTTGATCTCTACGACTGTTCAAAATTTGTTGAAGACCTTTAA 330  
730 CAAGTGGCTGAATTTGAGCAGAGTTCAAGAACTAGAAACCACTATACAGAAATTT 789  
329 CAAGTGGCTGAATTTGAGCAGAGTTCAAGAACTAGAAACCACTATACAGAAATTT 270  
790 TCTTTCACTGAGAACCTACTATATGAGGAATGAAACATCTGTTTGGGCCAAG 849  
269 TCTTTCACTGAGAACCTACTATATGAGGAATGAAACATCTGTTTGGGCCAAG 210  
850 AAACAGACCTCTGTTTGAAGCAATTAATTTTACCCCTTCAACCACTATGTC 509  
209 AAACAGACCTCTGTTTGAAGCAATTAATTTTATACCCCTTCAACCACTATGTC 150  
910 AACTAAAGATTTCTGTTGAGTCTCTTCAAAATTTTGAATGAGGATGTCGACAAAC 569

Db 149 AACTAAAGATTTCTGTTGAGTCTCTTCAAAATTTTGAATGAGGATGTCGACAAAC 90  
Qy 970 GTTCTATTTGTTTAAATTTGAAATATGTTTAACTATGAAATCCCTTTATTTA 1029  
Db 89 GTTCTATTTGTTTAAATTTGAAATATGTTTAACTATGAAATCCCTTTATTTA 30  
Qy 1030 AATACATATGAAAGAA 1046  
Db 29 AATACATTTGAAAGAA 13

RESULT 13  
AK032293  
LOCUS  
DEFINITION  
Mus musculus adult male olfactory brain cDNA, RIKEN full-length  
cDNA library, clone 643051BP20 product: similar to  
CERIOD-1/POFUSCINOSIS NEURONAL PROTEIN 5 (CLN5 PROTEIN) [Homo  
sapiens], full insert sequence.  
AK032293

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
AK032293.1 GI:26328112  
HTC; CAP trapper.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PubMed  
10349636  
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subcloning of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PubMed  
11042159  
3 Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
Komno, H., Akiyama, T., Nishi, K., Kibumai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PubMed  
11076861  
4 The RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PubMed  
11076861  
5 The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PubMed  
11076861  
6 (bases 1 to 2247)  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hitozane, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M.,  
Koya, S., Kurimura, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saito, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shitaki, T.,

## TITLE

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takahashi, A., S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M., and Hayashizaki, Y.  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

## COMMENT

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers

## FEATURES

source

```
1..2247
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6430516P20"
/db_xref="MGI:2395762"
/db_xref="taxon:10090"
/clone="6430516P20"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
<1..971
/note="unlabeled protein product; putative
similar to CERIOD-LIPOFUSCINOSIS NEURONAL PROTEIN 5 (CLN5
PROTEIN) [Homo sapiens] (SWISSPROT |075503, evidence:
FASTA, 73.9%ID, 86.7%length, match=1041)"
/codon_start=3
/protein_id="BAC27797.1"
/db_xref="GI:26328113"
/translation="LGLATLIGASPTSGQRPVYKHFSEPRTPDYCAKTECPTE
SPIPKMDNDVIEVLQAPIMEFKYGDLDGKLMHNDVGFSTLGNVTEWEL
FOJGNCETPHLRBPKSAPFMCNGAGCFEKGIDKMKNFVATLISGNTFENVA
EWKQDNRETCIYETETVAPRPGOAGTWEESYDGNFVRYTKLAERFTEKLET
NRYKFLYSGEPLIYLGNETSIFGPKAKTIALAKKPYGPRPYLSTKDFLMMFLKIF
DVTIHRQFLYFNFEXWFLPKRPFKITVEBTPLPTRHTTTDL"
```

## CDS

```
137 CGGGCTCGGGCATCCCTCCGCGCCGCACTGCGCGTGCCTTCAAGCGCTTGACTTC 196
21 CTGGGCGGCTCCCGACGCTCGGCAACGCTGCGCGGCATACAGCACTTCTCTTC 30
197 CGTCCAAAACCTGATCTTATTTGCAAGCTAAGTATCTTCTCCAACTGGCTCACT 256
81 CGTCCGAAGACAGATCCCTACTGTCACCTAAGTATCTTCTGCTCAACGCGCTCGCC 140
257 ATCCAGATTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 316
141 ATCCAGATTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 200
201 GAATTTAAATATGAGGACCTCTCGGACACTTTAACTTATGATGATGATGATGATG 260
317 GAATTTAAATATGAGGACCTCTCGGACACTTTAACTTATGATGATGATGATGATG 376
377 AGAATTAATTAATGAGGACCTCTCGGACACTTTAACTTATGATGATGATGATGATG 436
261 AGGACACACTGACAGGAGAAATACCAATATGATGATGATGATGATGATGATGATG 320
437 AACTGTAACTTCCCATCTCGACCTGAAATGATGATGATGATGATGATGATGATGATG 496
```

## ORIGIN

Query Match 61.7%; Score 668.6; DB 11; Length 2247;  
Best Local Similarity 81.6%; Pred. No. 8.8e-136;  
Matches 773; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 137 CGGGCTCGGGCATCCCTCCGCGCCGCACTGCGCGTGCCTTCAAGCGCTTGACTTC 196  
DB 21 CTGGGCGGCTCCCGACGCTCGGCAACGCTGCGCGGCATACAGCACTTCTCTTC 30  
QY 197 CGTCCAAAACCTGATCTTATTTGCAAGCTAAGTATCTTCTCCAACTGGCTCACT 256  
DB 81 CGTCCGAAGACAGATCCCTACTGTCACCTAAGTATCTTCTGCTCAACGCGCTCGCC 140  
QY 257 ATCCAGATTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 316  
DB 141 ATCCAGATTATGAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATG 200  
QY 201 GAATTTAAATATGAGGACCTCTCGGACACTTTAACTTATGATGATGATGATGATG 260  
DB 317 GAATTTAAATATGAGGACCTCTCGGACACTTTAACTTATGATGATGATGATGATG 376  
QY 377 AGAATTAATTAATGAGGACCTCTCGGACACTTTAACTTATGATGATGATGATGATG 436  
DB 261 AGGACACACTGACAGGAGAAATACCAATATGATGATGATGATGATGATGATGATG 320  
QY 437 AACTGTAACTTCCCATCTCGACCTGAAATGATGATGATGATGATGATGATGATGATG 496

DB 321 AACTGTAACTTCCCATCTCGACCTGAAATGATGATGATGATGATGATGATGATGATG 380  
QY 497 GCGGCGGCTCTTTTGAAGGAATGATGATGATGATGATGATGATGATGATGATGATG 556  
DB 381 GCGGCGGCTCTTTTGAAGGAATGATGATGATGATGATGATGATGATGATGATGATG 440  
QY 557 CAACTAGCAATATATGAGGAACATGTTCAACCAATGCGAAGTGGTGTAAACAGAC 616  
DB 441 GTCTTCAACATATCTCGGAAACATTTAAACAAAGTGGCGGTGATGATGATGATG 500  
QY 617 AATGAAACAGGAATTTATATGATGATGATGATGATGATGATGATGATGATGATG 676  
DB 501 AATGAAACAGGAATTTATGATGATGATGATGATGATGATGATGATGATGATG 560  
QY 677 GAGACATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 736  
DB 561 CAGACGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 620  
QY 737 GCTGATTTGAGGACAGATGATGATGATGATGATGATGATGATGATGATGATG 796  
DB 621 GCTGATTTGAGGACAGATGATGATGATGATGATGATGATGATGATGATGATG 680  
QY 797 AGTGAAGACCTACTTATCTGGAAGAAATGAAACATCTGTTTGGGCAACAGAAACAG 856  
DB 681 AGTGAAGACCTACTTATCTGGAAGAAATGAAACATCTTATGATGATGATGATGATG 740  
QY 857 ACTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 916  
DB 741 ACTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 800  
QY 917 GAATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 976  
DB 801 GAATTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 860  
QY 977 TTGTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1036  
DB 861 TTGTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 920  
QY 1037 TATGAAGAAATCCCTTATCTTACGAAACAAACCACTCTGCTGTTT 1083  
DB 921 TACGAAGAAATCCCTTATCTTACGAAACAAACCACTTATCGACTT 967

RESULT 14  
BX370641 mRNA linear EST 08-MAY-2003  
LOCUS BX370641 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL004YPO3 5-PRIME, mRNA sequence.  
ACCESSION BX370641  
VERSION BX370641.1 GI:30459737  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 908)  
AUTHORS La, W.B., Gruber, C., Jessee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7238.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAG053ZH03\_CS05016\_1&cluster=7238.f.  
Contact: Feng Liang Email: filiang@lifestech.com URL:  
http://fulllength.invitrogen.com/invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID: CS0BAG053ZH03\_CS05016\_1.  
Location/Qualifiers

## FEATURES

source

1. .908  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODL004P03"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

Query Match 59.5%; Score 645.4; DB 13; Length 908;  
Best Local Similarity 98.2%; Pred. No. 1e-130;  
Matches 663; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

## ORIGIN

410 GATGGTATGAACCTTTCCCACTTGGCACTGTATATTTCCCATCTCCGACCTGAATG 469  
1 GATATGCTGACCTTTTCCACTTGGCACTGTATATTTCCCATCTCCGACCTG-AAATG 59  
470 GATGCCCCCTTCTGATGATATCAAGGCGCTGCTTTTGGAGGAAATGATGATG 529  
60 GATGCCCCCTTCTGATGATATCAAGGCGCTGCTTTTGGAGGAAATGATGATG 119  
530 CACTGGAGAGAAATGGGACATTAAGTCAAGTATATTCAGAAACATGTTCAAC 589  
120 CACTGGAGAGAAATGGGACATTAAGTCAAGTATATTCAGAAACATGTTCAAC 179  
590 CAAATGGCAAAATGGGACATTAAGTCAAGTATATTCAGAAACATGTTCAAC 649  
180 CAAATGGCAAAATGGGACATTAAGTCAAGTATATTCAGAAACATGTTCAAC 239  
650 GATTAAGCAAGCCGCAAGAAAGGGGAGAGATGTTGATTCCTCAAGCTTTCCAAA 709  
240 GATTAAGCAAGCCGCAAGAAAGGGGAGAGATGTTGATTCCTCAAGCTTTCCAAA 299  
710 TTTGTTTAAAGACCTTTAACAAGTTGCTGAATTTGAGCAAGATTCAGAAACATGAA 769  
300 TTTGTTTAAAGACCTTTAACAAGTTGCTGAATTTGAGCAAGATTCAGAAACATGAA 359  
770 ACCAATATACAAAGATATTTCTTCACTGAGGAAACCTACTATCTGGGAAATGAA 829  
360 ACCAATATACAAAGATATTTCTTCACTGAGGAAACCTACTATCTGGGAAATGAA 419  
830 TCTGTTTGGGCAACAGAAACAGACTCTGTTGTTAGCAATAAAGATTTATAC 889  
420 TCTGTTTGGGCAACAGAAACAGACTCTGTTGTTAGCAATAAAGATTTATAC 479  
890 CCCTTCAACCAATTTGCAACTAAGAAATTTCTGTAATCTCTTGAATTTTGA 549  
480 CCCTTCAACCAATTTGCAACTAAGAAATTTCTGTAATCTCTTGAATTTTGA 539  
950 GCAATGATTTGCAACCAAGTTCTATTTGTTTAAATTTGAATTTGTTTAACT 1009  
540 GCAATGATTTGCAACCAAGTTCTATTTGTTTAAATTTGAATTTGTTTAACT 599  
1010 ATGAATTCCTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1069  
600 ATGAATTCCTCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 559  
1070 ACACTCTCTGTTTA 1084  
660 ACACTCTCTGTTTA 674

RESULT 15  
CB215456 641 bp mRNA linear EST 06-FEB-2003  
LOCUS CB215456  
DEFINITION NISC np06c07.y1 NICHD\_HS\_Utl1 Homo sapiens cDNA clone IMAGE:5937013  
5', mRNA sequence.

ACCESSION CB215456  
VERSION CB215456.1 GI:28263648  
KEYWORDS EST  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 641)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
CDNA Library Preparation: CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov  
Plate: LLM13163 row: E column: 14  
Seq primer: M13RP1 reverse primer (ABI).  
Location/Qualifiers  
1. .641  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5937013"  
/sex="female"  
/tissue\_type="normal endometrium, late proliferative phase, cycle day 13"  
/lab\_host="DH10B (Tl-resistance)"  
/clone\_lib="NICHD\_HS\_Utl1"  
/note="Organ: uterus; Vector: pCMV-Sport6.1.cdb (Resgen, Invitrogen Corporation); Site 1: NotI; Site 2: EcoRV; Cloned unidirectionally from microquantity amounts of mRNA from normal endometrial tissue (late proliferative phase, cycle day 13). Average insert size 1.9 kb. Library constructed by Resgen (Invitrogen Corporation)."

## FEATURES

source

## ORIGIN

Query Match 59.1%; Score 641; DB 14; Length 641;  
Best Local Similarity 100.0%; Pred. No. 9.3e-130;  
Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

96 CGCTGGCGTCTGTTGGCTCGCGGTGTTCCGGCTGTCCTCGGCAATCCCT 155  
1 CGCTGGCGTCTGTTGGCTCGCGGTGTTCCGGCTGTCCTCGGCAATCCCT 60  
156 CCGGCGGCACTGGCGGCTGCTCAAGCGCTTGAATTCGCTCCAAAACCTGATCCT 215  
61 CCGGCGGCACTGGCGGCTGCTCAAGCGCTTGAATTCGCTCCAAAACCTGATCCT 120  
216 ATGTCAAGCTAAGTATCTTCTGTCCAACTGGCTCACTATCCAGTTATGAGAGGTG 275  
121 ATGTCAAGCTAAGTATCTTCTGTCCAACTGGCTCACTATCCAGTTATGAGAGGTG 180  
276 ATGTCAAGCTAAGTATCTTCTGTCCAACTGGCTCACTATCCAGTTATGAGAGGTG 335  
181 ATGTCAAGCTAAGTATCTTCTGTCCAACTGGCTCACTATCCAGTTATGAGAGGTG 240  
336 TCGTGGGACACTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 395  
241 TCGTGGGACACTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 300  
396 AGAATCAACAATGGAATGATGAACTTTTCAACTGGCACTGTAATTTCCCATC 455  
301 AGAATCAACAATGGAATGATGAACTTTTCAACTGGCACTGTAATTTCCCATC 360  
456 TCGAATCAACAATGGAATGATGAACTTTTCAACTGGCACTGTAATTTCCCATC 515  
361 TCGAATCAACAATGGAATGATGAACTTTTCAACTGGCACTGTAATTTCCCATC 420

```
Qy 516 GAATGATGATGTTCACTGGAGGAAATGGACATTAGTTCAGTACACTATATCAG 575
Db 421 GAATGATGATGTTCACTGGAGGAAATGGACATTAGTTCAGTACACTATATCAG 480
Qy 576 GAAACATGTTCAACCAATGGCAAAGTGGGTGAACAGACAAATGAAACAGGAAATTATTT 635
Db 481 GAAACATGTTCAACCAATGGCAAAGTGGGTGAACAGACAAATGAAACAGGAAATTATTT 540
Qy 636 ATGAGACATGGAATGTAAAGCCAGCCCAAGAAAGGGGAGAGACATGTTGATTCCCT 695
Db 541 ATGAGACATGGAATGTAAAGCCAGCCCAAGAAAGGGGAGAGACATGTTGATTCCCT 600
Qy 696 ACGACTGTTCCAAATTGTTGTTAAGGACCTTAAACAAGTTG 736
Db 601 ACGACTGTTCCAAATTGTTGTTAAGGACCTTAAACAAGTTG 641
```

Search completed: April 22, 2004, 04:39:45  
Job time : 3276.24 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	1084	100.0	1486	9	US-09-122-883-1	Sequence 1, Appl
2	1084	100.0	1486	14	US-10-010-050A-1	Sequence 1, Appl
3	1067	98.4	1751	13	US-10-653-595-26	Sequence 26, Appl
4	1067	98.4	1751	13	US-09-397-945-26	Sequence 26, Appl
5	772.4	71.3	1038	9	US-09-122-883-13	Sequence 13, Appl
6	772.4	71.3	1038	14	US-10-010-050A-13	Sequence 13, Appl
7	491	45.3	494	9	US-09-864-761-2252	Sequence 2252, Appl
8	401.4	37.0	506	14	US-10-040-713-825	Sequence 825, Appl
9	380.2	35.1	474	9	US-09-864-761-583	Sequence 5832, Appl
10	266	24.5	497	16	US-10-864-049-1859	Sequence 1859, Appl
11	197.2	18.2	351	9	US-09-867-701-2579	Sequence 2579, Appl
12	133	12.3	473	9	US-09-954-456-1363	Sequence 1363, Appl
13	60	5.5	60	10	US-09-908-975-1033	Sequence 10339, Appl
14	45.4	4.2	1945	9	US-09-938-842A-5232	Sequence 5232, Appl

C	15	45.4	4.2	1945	11	US-09-938-842A-5232	Sequence 5332, App
C	16	44.4	4.1	1991	15	US-10-190-312A-45	Sequence 45, App
C	17	43.8	4.0	17419	15	US-10-239-676-100	Sequence 100, App
C	18	43.8	4.0	17419	15	US-10-311-455-1268	Sequence 1268, App
C	19	43.8	4.0	17419	15	US-10-240-453-112	Sequence 112, App
C	20	43.2	4.0	6361	15	US-10-311-455-1113	Sequence 1113, App
C	21	43.2	4.0	3673778	15	US-10-312-841-2	Sequence 2, App
C	22	42.6	3.9	6357	15	US-10-311-455-70	Sequence 70, App
C	23	42.4	3.9	606	13	US-10-424-599-99310	Sequence 99310, A
C	24	42.4	3.9	2235	15	US-10-156-761-4003	Sequence 4003, App
C	25	42.4	3.9	6283	15	US-10-311-455-62	Sequence 62, App
C	26	42.4	3.9	7072	13	US-10-221-613-347	Sequence 347, App
C	27	42.4	3.9	9025508	15	US-10-156-761-1	Sequence 1, App
C	28	42.2	3.9	43058	9	US-09-954-456-292	Sequence 292, App
C	29	42.2	3.9	43058	9	US-09-954-456-529	Sequence 529, App
C	30	42.2	3.9	43058	9	US-09-880-107-3950	Sequence 3950, App
C	31	42.2	3.9	167343	9	US-09-962-436-281	Sequence 281, App
C	32	42	3.9	167343	9	US-09-962-436-281	Sequence 273, App
C	33	41.8	3.9	479	10	US-09-864-824R-273	Sequence 21958, A
C	34	41.6	3.8	9254	13	US-10-221-714A-113	Sequence 113, App
C	35	41.6	3.8	9254	15	US-10-239-676-85	Sequence 85, App
C	36	41.6	3.8	9254	15	US-10-240-453-91	Sequence 91, App
C	37	41.4	3.8	3156	15	US-10-156-761-833	Sequence 833, App
C	38	41.4	3.8	9025608	15	US-10-156-761-1	Sequence 1, App
C	39	41.2	3.8	467	10	US-09-918-999-13111	Sequence 13111, A
C	40	41.2	3.8	640681	9	US-09-790-988-1	Sequence 1, App
C	41	41.2	3.8	3673778	15	US-10-312-841-1	Sequence 108, App
C	42	41	3.8	12781	13	US-10-221-714A-108	Sequence 18, App
C	43	40.8	3.8	5979	15	US-10-239-676-18	Sequence 26, App
C	44	40.8	3.8	5979	15	US-10-240-453-26	Sequence 26, App
C	45	40.8	3.8	12606	9	US-09-957-974-2	Sequence 2, App

## ALIGNMENTS

```

RESULT 1
US-09-122-383-1
: Sequence 1, Application US/09122383A
: Patient No. US20020040092A1
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Gilbertson, Dextra G.
: TITLE OF INVENTION: SECRETED PROTEINS ENCODED
: TITLE OF INVENTION: CHROMOSOME 13
: FILE REFERENCE: 97-38
: CURRENT APPLICATION NUMBER: US/09/122,383A
: CURRENT FILING DATE: 1998-07-24
: EARLIER APPLICATION NUMBER: 60/053,613
: EARLIER FILING DATE: 1997-07-24
: NUMBER OF SEQ. ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 1486
: TYPE: DNA
: ORGANISM: Homo sapien
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (47)...(1084)
: US-09-122-383-1

```

Query Match	100.0%	Score 1084;	DB 9;	Length 1486;
Best Local Similarity	100.0%	Pred. No. 7e-283;		
Matches 1084; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

[illegible]





```

661 CCCAGAAAAGGGGCGAGAGACATGGTTGATTCCTACGACTGCTCCAAATTGTGTAG 720
721 GACCTTTAAACAAGTGGCTGATTTGGAGCAGAGTCAAGACATAGAAACCACTATAC 780
721 GACCTTTAAACAAGTGGCTGATTTGGAGCAGAGTCAAGACATAGAAACCACTATAC 780
781 AAGATATTTCTTTCATGAGTGAACCTTATCTGGAATGAAACATCTGTTTTGG 843
781 AAGATATTTCTTTCATGAGTGAACCTTATCTGGAATGAAACATCTGTTTTGG 843
841 GCCAAGAGAAACAAGACTTGTGTTAGCCATATAAAGATTTATACCCCTTGAACC 903
841 GCCAAGAGAAACAAGACTTGTGTTAGCCATATAAAGATTTATACCCCTTGAACC 900
901 ACATTTGCCAATTAAGAAATTTCTGTGAGTCTCTGCAATTTTGATGACATGTG 960
901 ACATTTGCCAATTAAGAAATTTCTGTGAGTCTCTGCAATTTTGATGACATGTG 960
961 GCACAAACAGTCTATTTGTTTATATTTGAAATTTGTTTATCTATGAAATTTCC 1020
961 GCACAAACAGTCTATTTGTTTATATTTGAAATTTGTTTATCTATGAAATTTCC 1020
1021 TTTTATTAATAATACATATGAGAAATCCCTTATCTATGAGAAACAAACATCTCTG 1080
1021 TTTTATTAATAATACATATGAGAAATCCCTTATCTATGAGAAACAAACATCTCTG 1080
1081 TTTT 1084
1081 TTTT 1084

```

## RESULT 3

JS-10-653-595-26  
 Sequence 26, Application US/10653595  
 Publication No. US20040048304A1

## GENERAL INFORMATION:

APPLICANT: Ruben et. al.

TITLE OF INVENTION: 95 Human secreted proteins

FILE REFERENCE: P2027P1C1

CURRENT FILING DATE: 2003-09-03

PRIOR FILING DATE: 1999-09-17

PRIOR FILING DATE: 1999-09-17

PRIOR FILING DATE: 1999-03-18

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

PRIOR FILING DATE: 1998-03-19

```

NAME/KEY: SITE
LOCATION: (1557)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1689)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1729)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1735)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: SITE
LOCATION: (1741)
OTHER INFORMATION: n equals a,t,g, or c
US-10-653-595-26

```

## Query Match

Best Local Similarity 98.4%; Score 1067; DB 13; Length 1751; Pred. No. 3.2e-278; Indels 0; Gaps 0; Matches 1067; Conservative 2; Mismatches 2;

```

14 GGGGAGAGGTAGACACGGGAGAGGCGGAGATGCGGCGGCGGCGGCGGCGGCGG 73
16 GCGCAGAGGTAGACACGGGAGAGGCGGAGATGCGGCGGCGGCGGCGGCGGCGG 75
74 GGAAGGCTTCTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 133
76 GGAAGGCTTCTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 135
134 TCCGCGGCTTCTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193
136 TCCGCGGCTTCTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 195
194 TCCGCGGCTTCTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253
196 TCCGCGGCTTCTGAGTGGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 255
254 CCGATCCAGTTATGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 313
256 CCGATCCAGTTATGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 315
314 TGGGATTTAAATATGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 373
316 TGGGATTTAAATATGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 375
374 TCCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 433
376 TCCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 435
434 GGCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 493
436 GGCACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 495
494 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 553
496 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 555
554 GTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 613
556 GTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 615
614 GACATGAAACAGGAATTTATATGAGATGATGATGATGATGATGATGATGATGATG 673
616 GACATGAAACAGGAATTTATATGAGATGATGATGATGATGATGATGATGATGATG 675
674 GCGAGACATGTTGATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 733
676 GCGAGACATGTTGATTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 735

```

2y	734	TTGGCTGAATTTGGAGCAGAGTTCAAGACATGGAACCAACTATACAAAGAAATTTCTT	793
2y	736	TTGGCTGAATTTGGAGCAGAGTTCAAGACATGGAACCAACTATACAGGATATTTCTT	795
2y	794	TACAGTGAACACCTACTATATCGGAGAAATGAAACATCTGTTTTGGGCCAACAGAAAC	853
2y	796	TACAGTGAACACCTACTATATCTGGGAAATGAAACATCTGTTTTGGGCCAACAGAAAC	855
2y	854	AAGACTCTTGTTAGCCATAAAAGATTTATTACCCCTTCAACACATTTGCCAACT	913
2y	856	AAGACTCTTGTTAGCCATAAAAGATTTATTACCCCTTCAACACATTTGCCAACT	915
Qy	914	AAAGAAATTTCTGTGAGTCTCTTGCAAAATTTTATGACAGATGTGCAAAACAGTTC	973
Db	916	AAAGAAATTTCTGTGAGTCTCTTGCAAAATTTTATGACAGATGTGCAAAACAGTTC	975
Qy	974	TATTTGTTTATATTTTGAATATATGGTTTTTACCTATGAATTCCTTTATTTAAATA	1033
Db	976	TATTTGTTTATATTTTGAATATATGGTTTTTACCTATGAATTCCTTTATTTAAATA	1035
Qy	1034	ACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGTGTTTA	1084
Db	1036	ACATATGAAGAAATCCCTTACCTATCAGAAACAAACACTCTCTGTGTTTA	1086

RESULT 4  
INC-09-39

```

/ Sequence 26, Application US/09397945
/ US09-397,945
/ Publication No. US2003006519A1
/ GENERAL INFORMATION:
/ APPLICANT: Human Genome Sciences, Inc. et al.
/ TITLE OF INVENTION: 95 Human secreted proteins
/ FILE REFERENCE: P2027P1
/ CURRENT APPLICATION NUMBER: US/09/397,945
/ CURRENT FILING DATE: 1999-09-17
/ PRIOR APPLICATION NUMBER: PCT/US99/05804
/ PRIOR FILING DATE: 1999-03-18
/ PRIOR APPLICATION NUMBER: 60/078,566
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/078,576
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/078,573
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/078,574
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/078,579
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/080,314
/ PRIOR FILING DATE: 1998-04-01
/ PRIOR APPLICATION NUMBER: 60/080,312
/ PRIOR FILING DATE: 1998-04-01
/ PRIOR APPLICATION NUMBER: 60/078,578
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/078,581
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/078,577
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/078,563
/ PRIOR FILING DATE: 1998-03-19
/ PRIOR APPLICATION NUMBER: 60/080,313
/ PRIOR FILING DATE: 1998-04-01
/ NUMBER OF SEQ ID NOS: 470
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 26
/
/ LENGTH: 1751
/
/ TYPE: DNA
/
/ ORGANISM: Homo sapiens
/
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (1520)
/ OTHER INFORMATION: n equals a,t,g, or c
/ NAME/KEY: SITE
/ LOCATION: (1557)
/

```

```

: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (1689)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (1729)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (1735)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: (1741)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-397-945-26

```

Query Match	98.4%	Score 1067	DB 13	Length 1751
-------------	-------	------------	-------	-------------

Best Local Similarity	99.6%;	Pred. No. 3.2e-278;	
Matches 1067;	Conservative	2;	Mismatches
			Indels
			Gaps
			0.7

QY	14	GGGCAAGAGGTTAGACACAGGCAAGGGGCGCGAGATGAGGAGGGGGGCGGGGCGGAGCTTCGG	73
Db	16	GCGCAAGAGAGGTAGACACGGCAAGGGGCGCGAGATGAGGAGGGGGCGGGGCGGAGCTTCGG	75
QY	74	GGAGCGCTTCCTGAGTGTGGAGCGCTGGCGCTGTCTGGCTCGCGGTGTTCGGGCTGG	133
Db	76	GGAGCGCTTCCTGAGTGTGGAGCGCTGGCGCTGTCTGGCTCGCGGTGTTCGGGCTGG	135
QY	134	TCCGGGCTCTGGGAGTCCCTCCCGGCGCACTGGCCGGTGCCCTACAGCGCTTTAGC	193
Db	136	TCCGGGCTCTGGGAGTCCCTCCCGGCGCACTGGCCGGTGCCCTACAGCGCTTTAGC	195
QY	194	TTCCGTCGCAAAACCTGATCCCTTATTTGTCAACCTTAAGTATACCTTTGTCGCAACTGGCTCA	253
Db	196	TTCCGTCGCAAAACCTGATCCCTTATTTGTCAACCTTAAGTATACCTTTGTCGCAACTGGCTCA	255
QY	254	CCTATCCCAAGTTATGGAAGGATGATGATGACATTGAAGTTTTCGATTCAAGCCCCAGTA	313
Db	256	CCTATCCCAAGTTATGGAAGGATGATGATGACATTGAAGTTTTCGATTCAAGCCCCAGTA	315
QY	314	TGGGAATTTAAATATGAGAGCTCCTCGGAGCACTTGAAATTTATGATGATGCCATTGGA	373
Db	316	TGGGAATTTAAATATGAGAGCTCCTCGGAGCACTTGAAATTTATGATGATGCCATTGGA	375
QY	374	TTCAAGAGTCACTTAATCTGGCAAGACTACACATGGAATGGTATGAAGCTTTTCCAACT	433
Db	376	TTCAAGAGTCACTTAATCTGGCAAGACTACACATGGAATGGTATGAAGCTTTTCCAACT	435
QY	434	GGCAACTGTACATTTCCCATCTCCGACCTGAAATGGAGCCCTTCTGCTGTATCAA	493
Db	436	GGCAACTGTACATTTCCCATCTCCGACCTGAAATGGAGCCCTTCTGCTGTATCAA	495
QY	494	GGGCGTGCCTGCTTTTGTGAGGAAATTGATGATGTTCATCGAAGGAAATGGGACATTA	553
Db	496	GGGCGTGCCTGCTTTTGTGAGGAAATTGATGATGTTCATCGAAGGAAATGGGACATTA	555
QY	554	GTTCAAGTACATATATCAGAAACATGTTCAACCAATGGCAAAAGTGGTGAACAG	613
Db	556	GTTCAAGTACATATATCAGAAACATGTTCAACCAATGGCAAAAGTGGTGAACAG	615
QY	614	GACATGAAACAGGAATTTATATGAGACATGGAATGTAAAGCCAGCCAGAAAAGGGG	673
Db	616	GACATGAAACAGGAATTTATATGAGACATGGAATGTAAAGCCAGCCAGAAAAGGGG	675
QY	674	GCAGAGCAATGTTGATCTCTACGACTGTTCAAAATTTGTGTAAAGACTTTAACAG	733
Db	676	GCAGAGCAATGTTGATCTCTACGACTGTTCAAAATTTGTGTAAAGACTTTAACAG	735
QY	734	TTGAGTGAATTTGAGAGAGTTCAAGACATAGAAACCAACTATACAGAAATTTCTT	793
Db	736	TTGAGTGAATTTGAGAGAGTTCAAGACATAGAAACCAACTATACAGAAATTTCTT	795
QY	794	TACAGTGAACCTACTTATCTGGGAAATGAACATCTGTTTTGGGCCAACAGGAAC	853